CRISP: A Computational Model of Fixation Durations in Scene Viewing

Antje Nuthmann (1), Tim J. Smith (1), Ralf Engbert (2), John M. Henderson (1)

(1) University of Edinburgh, UK

(2) University of Potsdam, Germany

Running Head: Modeling Fixation Durations in Scenes

Keywords: computational modeling; eye movements; naturalistic scenes; fixation durations; onset delay

Corresponding Author:

Antje Nuthmann

University of Edinburgh

Psychology Department

7 George Square

Edinburgh, EH8 9JZ, UK

E-Mail: Antje.Nuthmann@ed.ac.uk

Phone: +44 (0)131 650 3459

Abstract

Eye-movement control during scene viewing can be represented as a series of individual decisions about where and when to move the eyes. While substantial behavioral and computational research has been devoted to investigating the placement of fixations in scenes, relatively little is known about the mechanisms that control fixation durations. Here we propose a computational model (CRISP) that accounts for saccade timing and programming and thus for variations in fixation durations in scene viewing. First, timing signals are modeled as continuous-time random walks. Second, difficulties at the level of visual and cognitive processing can inhibit and thus modulate saccade timing. Inhibition generates moment-by-moment changes in the random walk's transition rate and processing-related saccade cancellation. Third, saccade programming is completed in two stages: an initial, labile stage that is subject to cancellation, and a subsequent, non-labile stage. Several simulation studies tested the model's adequacy and generality. An initial simulation study explored the role of cognitive factors in scene viewing by examining how fixation durations differ under different viewing task instructions. Additional simulations investigated the degree to which fixation durations are under direct moment-to-moment control of the current visual scene. The present work further supports the conclusion that fixation durations, to a certain degree, reflect perceptual and cognitive activity in scene viewing. Computational model simulations contribute to our understanding of the underlying processes of gaze control.

224 words

Introduction

Looking at a scene typically involves active visual sampling, with observers moving their eyes approximately three times each second via saccadic eye movements (Buswell, 1935; Yarbus, 1967; for recent reviews see Henderson, 2003; Henderson, 2007; Rayner, 1998, 2009b). Two important aspects of gaze control during scene perception are where fixations are directed (fixation position), and how long they typically remain in a given location (fixation duration). Current computational models of gaze control in scenes exclusively consider the "where" decision; they predict fixation locations but do not account for fixation durations (see Henderson, 2003; Rayner, 2009b). Furthermore, current models do not attempt to explain the mechanisms underlying saccade programming and thus ignore the restrictions that arise from the operation of the oculomotor system. The present paper begins to address this shortcoming by introducing a timer (C)ontrolled (R)andom-walk with (I)nhibition for (S)accade (P)lanning model. CRISP is a theoretical approach and a computational model that, based on minimal assumptions, accounts for variations in fixation durations during scene viewing.

Where viewers fixate in a scene is an overt behavioral manifestation of selective attention. Therefore, most studies of attention allocation during scene viewing focus on explaining the processes that determine fixation location rather than fixation duration (e.g., Baddeley & Tatler, 2006; Henderson, Malcolm, & Schandl, 2009; Itti & Koch, 2000; Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Tatler, Baddeley, & Vincent, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006; Underwood & Foulsham, 2006). Theoretical models of visual attention allocation can be contrasted as primarily bottom-up versus top-down, with these views forming the extremes of a continuum (see Henderson,

2003; Henderson, Brockmole, Castelhano, & Mack, 2007, for discussion). The dominant view in the computational vision literature has been the visual saliency hypothesis according to which bottom-up stimulus-based information generated from the image drives the allocation of visual attention and thus the placement of fixations in a scene (e.g., Itti & Koch, 2000, 2001; Itti, Koch, & Niebur, 1998). A model combining both bottom-up and top-down information is the Contextual Guidance model for object search in real-world scenes (Torralba et al., 2006). In this model, fixation selection is driven from a bottom-up salience computation modulated by contextual priors or task constraints. More recent proposals have suggested that top-down guidance plays the primary role in eye guidance and replace the concept of salience with an unprioritized input representation (Henderson et al., 2007; Henderson et al., 2009).

When people inspect a scene, they not only generate considerable variability in fixation locations but also in fixation durations. The average fixation duration during scene viewing is about 300 ms, but there is substantial variability around this mean both within an individual and across individuals (Castelhano & Henderson, 2008; Henderson & Hollingworth, 1998; Rayner, 1998, 2009b) (see also Figure 3). In the literature on human gaze control, the influence of visual and cognitive factors on fixation duration is widely acknowledged. In reading, such relationships have been extensively investigated over the past 30 years and are now explicitly incorporated within computational models of eye-movement control (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Rayner, & Pollatsek, 2003). Fixation durations in reading are influenced by low-level nonlinguistic factors (e.g., word length, within-word fixation position) but also by various lexical, syntactic, and discourse factors (see Rayner, 1998, for a review). As an example,

fixation durations are shorter on words that are highly predictable from the context of the narrative (e.g., Balota, Pollatsek, & Rayner, 1985).

In the scene perception literature, however, the relationship between fixation durations and visual-perceptual and cognitive influences is less well established. A compelling demonstration of top-down influences on eye movements in scene viewing originates from task effects. While early studies highlighted the role of viewing task on the choice of fixation locations (Buswell, 1935; Yarbus, 1967), recent studies have also considered task effects (such as searching for an object in the scene or memorizing the scene) on fixation durations. Individual fixation durations (i.e., the duration of each discrete fixation) are typically found to be shorter during visual search than memorization (Henderson, Weeks, & Hollingworth, 1999; Võ & Henderson, 2009; but see Castelhano, Mack, & Henderson, 2009). Furthermore, there is some evidence that factors associated with the currently fixated scene region affect fixation duration. Specifically, fixation durations are sensitive to global image degradation. For example, reducing the luminance of a scene leads to increased fixation duration (Loftus, 1985; Loftus, Kaufman, Nishimoto, & Ruthruff, 1992). Mannan, Ruddock, and Wooding (1995) showed that low-pass filtered images produce longer fixation durations than either high-pass or unfiltered images. First-pass gaze durations[1] are also influenced by object and scene semantics, with longer gaze durations on semantically informative (i.e., less consistent) than uninformative (i.e., more consistent) objects (De Graef, Christiaens, & D'Ydewalle, 1990; Friedman, 1979; Henderson et al., 1999; Hollingworth, Williams, & Henderson, 2001; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Võ & Henderson, 2009). Whether semantic informativeness also affects individual first fixation durations (rather

than the sum of durations) on an object or a region in a scene is less clear (see Henderson & Hollingworth, 1998, for discussion). In sum, there is preliminary evidence that some of the observed variability in individual fixation durations is controlled by visual and cognitive factors associated with the currently fixated scene region. We will term this evidence for *direct control*.

A powerful paradigm for testing whether fixation duration is under direct control of the current stimulus is the *stimulus onset delay paradigm*. During the saccade prior to a pre-specified critical fixation, the stimulus (e.g., a text, a scene, or a visual search display) is removed from view. Following a manipulated delay period, the stimulus reappears. If fixation durations are directly controlled by the current visual input, then the critical fixation durations should systematically increase with the delay. A now classic study applied the stimulus onset delay paradigm to text reading (Morrison, 1984; see also Rayner & Pollatsek, 1981). Fixation durations showed as two populations. One population indeed increased with stimulus onset delay, indicating that fixation duration is under direct control. There was, however, a second population of fixations that ended while the mask was still present, suggesting that their preparation already started on the previous fixation. Morrison (1984) interpreted the results in terms of partially parallel saccade programming. This notion is part of Morrison's qualitative model of eye-movement control in reading where sequential shifts of attention from one word to the next form the link between word recognition and eye-movement control. The sequential attention shift model underwent several modifications and extensions (e.g., Henderson & Ferreira, 1990), which culminated in the development of the E-Z Reader model as the first fully implemented model of eye-movement control in reading (e.g., Reichle,

Pollatsek, Fisher, & Rayner, 1998; Reichle et al., 2003).

Recently, a series of experiments utilized the stimulus onset delay paradigm to investigate the control of individual fixation durations in scene viewing (Henderson & Pierce, 2008; Henderson & Smith, 2009; see also Shioiri, 1993; van Diepen, Wampers, & d'Ydewalle, 1998). In the *scene onset delay* (SOD) paradigm, participants examine photographs of real-world scenes while engaged in a viewing task (e.g., scene memorization, visual search). During the saccade just prior to a pre-specified critical fixation, the entire scene is replaced with a mask, which delays the onset of the scene. Following the delay period, which is variable in time, the scene reappears. If there is direct control of the critical fixation duration, one would expect programming of the eye movement to be delayed until there is scene information present on which to base the programming. The results from these experiments consistently demonstrate that fixation durations in scene viewing comprise two fixation populations. One population increases with scene onset delay, whereas the second population remains relatively constant across delays. The qualitative pattern of results is consistent with the results from studies of reading. The scene onset delay studies have allowed different theoretical ideas about the control of fixation durations during scene viewing to be contrasted (Henderson & Pierce, 2008; Henderson & Smith, 2009).

Present Work

The aim of the present work was to develop a theoretical framework and a computational model that accounts for variations in fixation durations in scene viewing. We first introduce the basic architecture of the CRISP model. This is followed by simulations with the baseline model. The model's adequacy and generality is examined

by three further simulation studies. In particular, model assumptions are tested using data from (1) a viewing task manipulation, (2) a scene onset delay paradigm, and (3) a mask onset delay paradigm. We note that the model, in its current implementation, does not perform an analysis of scene content. Although this is a limitation, the key point of the paper is to provide a general computational framework for exploring the extent to which fixation durations are under perceptual and cognitive control during scene viewing. We will demonstrate that investigating global effects of viewing task and/or effects of global scene processing difficulty on the control of fixation duration is a suitable starting point for this endeavor.

Basic Model Architecture

The present work is motivated by the idea that eye movements in scene perception represent a special case of a more general theory of eye movements and action. In particular, our modeling efforts draw inspiration from a group of models of eye-movement control sharing certain core assumptions on saccade timing (Engbert et al., 2005; Findlay & Walker, 1999; Yang & McConkie, 2001). Findlay and Walker (1999) proposed a very general qualitative model of saccade generation, based on parallel processing in "when" and "where" streams and inhibition. The model is built on the neurophysiologically motivated assumption of two separate pathways concerned with the spatial ("where") and the temporal ("when") programming of eye movements. Findlay and Walker proposed that the timing of saccade initiation is determined mainly by the "conflict resolution" competitive push-pull interaction between a *fixation centre* and a *move centre*. In their Competition-Inhibition Theory, Yang & McConkie (2001) extended Findlay & Walker's (1999) framework to account for fixation duration distributions

observed in reading (see Yang, 2006, for an implementation). The SWIFT model of saccade generation (Engbert, Longtin, & Kliegl, 2002; Engbert et al., 2005) represents a dynamical systems approach to eye-movement control in reading. For the "when" decisions, SWIFT postulates a *random saccade timer*, which can be modulated by the cognitive load imposed by foveal visual and lexical processing.[2]

Here, we propose and implement a random-walk process with inhibition plus two-stage saccade programming model to account for fixation durations in scene viewing. First, a random walk process generates inter-saccadic intervals and thus variations in fixation durations. Second, we assume that difficulties in moment-by-moment visual and cognitive processing immediately inhibit (i.e., delay) saccade initiation, essentially leading to longer fixation durations. The random-walk architecture of the saccade timer allows for a continuous crosstalk between saccade preparation and visual-cognitive processing. Third, saccade programming is completed in two stages: an initial, labile stage that is subject to cancellation, and an ensuing, non-labile stage (Becker & Jürgens, 1979; Reichle et al., 1998). In the following, these three modeling principles will be discussed in detail.

Rhythmic Timing

In the CRISP model, a rhythmic saccade timer generates variations in fixation durations (Figure 1). The underlying intuition is that we initiate saccade programs according to some preferred mean rate (Engbert et al., 2002; Engbert et al., 2005). There is growing support for the concept that the control of motor activity by the central nervous system may be autonomous (Engbert & Kliegl, 2001) or rhythmic in nature (see McAuley, Rothwell, & Marsden, 1999, for discussion). As far as the oculomotor system

is concerned, in a study by McAuley et al. (1999) anticipatory eye movements were generated by having participants track an intermittently obscured sinusoidally moving target. The results suggest that rhythmic activity originating in the central nervous system may modulate human eye movements.

Insert Figure 1 about here

In CRISP, the saccade timer is implemented as a random walk process. This basic idea is shared by a group of information accumulator models. These models come in many different flavors and have a long and successful history in accounting for data in simple manual or saccadic decision tasks (see Smith & Ratcliff, 2004, for a review and taxonomy). Note that, in our model, the random walk process controls the generation of intervals between two decisions to start the programming of new saccades, not saccade execution.[3] The random walk timing signal accumulates towards a positive response boundary (i.e., a threshold). Once the threshold is reached, a new saccade program is initiated (see below). The random walk timing signal creates a trajectory over time, which can be modulated at any point by visual-cognitive events (Figure 1). As a consequence, inhibitory influences from the visual-cognitive module will delay the point in time when the threshold is reached and a new saccade program is triggered. The threshold is constant and the variability in timing signals arises from randomness in the rate of growth of the accumulator (cf., Carpenter & Williams, 1995; Ratcliff & Rouder, 1998). There is neurophysiological evidence in favor of such a fixed-threshold, variable-growth architecture and against an alternative variable-threshold, fixed-growth architecture

(Hanes & Schall, 1996).

Implementation of the Random Walk Process

In CRISP, saccade timing is implemented as a discrete-state continuous-time random walk process with exponentially distributed waiting times between elementary transitions. The main parameter is the transition rate for the random walk's increments (i.e., the elementary jumps toward threshold), which is modulated by visual-cognitive processing (Figure 1) and determines how fast the process of saccade timing operates. We define the transition rate r_1 as

$$r_1 = \frac{N}{t_{sac}}, \tag{1}$$

where N is the number of states the process can adopt, and t_{sac} is the mean duration of the timing signal. We assume that the transition probability from the current state to the next state depends on the past only through the current state, so that the random walk introduced here falls into the broad category of *Markov processes* (Gardiner, 2004). As an efficient method for the numerical simulation of Markov processes we use the *minimal process method* (Gillespie, 1976, 1978), which represents an exact simulation procedure. This algorithm has been applied to theoretical models in areas as diverse as physiology (Fricke & Schnakenberg, 1991), molecular biology (Elowitz & Leibler, 2000), and reading (Engbert & Kliegl, 2003). Let us consider a system in state S_m, having arrived there at time t, which performs the next step to an adjoined state S_n at time $t + \tau$. The *waiting time* τ is defined as the time interval to the next transition. For a Markov process with N adjoined states to the current state S_m, the probability distribution $\rho(\tau)$ of the waiting time τ follows an exponential function,

$$\rho(\tau) = W_m e^{-W_m t} \quad \text{with} \quad W_m = \sum_{n \neq m} W_{nm}, \tag{2}$$

where W_m is the total transition probability, i.e., the sum of transition probabilities W_{nm} to all adjoining states S_n (i.e., all states to which the process can make a transition). The random walk process implemented here is a one-step process with a single possible transition $m \mapsto m+1$ from a given state S_m to its adjoined state S_{m+1}. In this case, all components of the matrix of transition probabilities W_{nm} are zero except for $W_{m+1,m} = r_1$, i.e., $W_n = W_{m+1,m} = r_1$. The mean waiting time in a given state S_m is given by the inverse of the total transition probability, i.e.,

$$\langle \tau \rangle = \frac{1}{W_m} = \frac{1}{w_0} \tag{3}$$

Since the waiting time is an exponentially distributed variable, it can be transformed from a uniformly distributed random number by

$$\tau = -\frac{1}{w_0} \log(1 - r), \tag{4}$$

where r is a random number with equal probability in $0 \leq r < 1$ (cf., Gillespie, 1978). It is important to note that the proportion between standard deviation σ_1 and mean μ_1 of the saccade timer is monotonically related to the number of states N of the random walk process by

$$\frac{\sigma_1}{\mu_1} = \frac{1}{\sqrt{N}} \tag{5}$$

Thus, the value of N, i.e., the threshold of the random walk process, is a free parameter for adjusting the noise level of the saccade timer (see Appendix A). A high N leads to small variance and thus deterministic behavior.

Inhibition: Processing Demands Modulate the Random Walk Timing Process

We further assume that processing difficulty can inhibit and thus modulate saccade timing and programming. According to the model architecture, the visual-cognitive processing module can (1) inhibit the random walk's transition rate and/or (2) cancel labile saccade programs (inhibitory elements are marked in red in Figure 1). Conceptualizing the timer as a random walk process allows for a continuous crosstalk between visual-cognitive processing and saccade timing: The random walk creates a trajectory over time, which can be modulated at any point by visual-cognitive events. Generally put, as processing demands increase the processing speed decreases, which will slow down the random walk saccade timer. Intuitively, greater processing difficulty increases the time needed to reach the completion threshold. This delays the initiation of a new saccade program, and eventually leads to longer fixation durations.

Two-Stage Saccade Programming

Once the implemented random walk process reaches threshold, a new saccade program is initiated. We assume that saccade programming is completed in two stages: an initial, *labile* stage that is subject to cancellation, and an ensuing, *non-labile* stage in which the program can no longer be cancelled (Becker & Jürgens, 1979; adopted by Reichle *et al.*, 1998; Findlay & Walker, 1999; Engbert *et al.*, 2005). Evidence for these mechanisms is provided by a series of experiments by Becker and Jürgens (1979) who used a much simpler situation than scene viewing. In the *double step paradigm* a target makes two successive movements in a quick sequence. Participants are asked to follow the target with their eyes. The paradigm allows investigation of how saccade initiation is controlled and how saccade targets are selected by measuring the effects of the second stimulus step on the programming of the saccade to the first step. The independent

variable is the time *D* elapsing between the initial saccade and the second target step. Participants' performance is stochastic and, depending on *D*, shows three principal outcomes. First, if *D* is *short* (i.e., when the two shifts of the target are close to one another in time), participants make only one eye movement to the second, *final* location of the target. Here, the oculomotor system begins programming a saccade to the second target location while the saccade program related to the first target location is still in its labile stage of development. In such a case, the first program is cancelled and only the second program is executed. Second, when *D* is *long* (i.e., when the two steps are relatively far apart in time) participants make two saccades. The first saccade is directed to the first shifted target location and the second to the second shifted location, with the first of these fixations being relatively short (about 50 ms). Here, the second saccade program is initiated while the first program is already in its non-labile stage and can no longer be altered. In such a case, both saccades will be executed, which typically results in a short fixation between the two saccades. Third, an *intermediate* range of *D* values (i.e., if the time between the two shifts of the target were intermediate) results in compromise saccades with endpoints landing between the two positions occupied by the target. Becker and Jürgens (1979) proposed that double step saccade performance could be understood as the outcome of a race between the processes producing saccades to the initial and final target locations (see Camalier et al., 2007, for an implementation of such a race model, and Ludwig, Mildinhall, & Gilchrist, 2007, for an alternative stochastic accumulator model).

Using a saccade stop-signal task, recent work has examined the neural basis of saccade programming stages (see Schall, 2004, for a review). Notably, certain neurons in

the frontal eye field and superior colliculus show different activation when saccades are executed versus cancelled, and these differences in activation occur within the time that the saccade is cancelled (Hanes, Patterson, & Schall, 1998; Paré & Hanes, 2003). During saccade preparation, neurons with movement-related activity begin to increase toward a saccade execution trigger threshold. When partially prepared saccades are cancelled, these neurons fail to reach the threshold activation level. Instead, the saccade-related activity decreases rapidly after the stop signal is presented.

In sum, existing research suggests that saccade programming is completed in different stages and that saccade programs can partly overlap in time (see also Findlay, Brown, & Gilchrist, 2001; McPeek, Skavenski, & Nakayama, 2000; Walker & McSorley, 2006). CRISP incorporates these principles (see also Engbert et al., 2005; Findlay & Walker, 1999; Reichle et al., 1998). Figure 2 visualizes how random walk timing and the saccade-programming module work together. Once the random walk process reaches threshold, a new saccade program is initiated. It enters a preliminary *labile* stage with an average duration τ_{lab}. If another saccade program is initiated during the labile stage, the first program will be cancelled and only the second program will be executed. Thus, the new labile saccade program will always override an old one, if there is one (see Figure 2 for an example). At the end of the labile stage, a point of no return is reached and the saccade can no longer be cancelled – the program is now *non-labile* (average duration τ_{nlab}). Consequently, if a second program is initiated during the non-labile stage of the first one, both will be executed. At the end of the non-labile saccade programming stage, the saccade is *executed* (average duration τ_{ex}).

Insert Figure 2 about here

Generation of Fixation Durations

When viewing a scene, our eyes alternate between fixations and saccades. Saccade latency is defined as the time needed to program an eye movement. In the model, we derive saccade latency l_{sac} as the sum of the implemented saccade programming stages, i.e.,

$$l_{sac} = \tau_{lab} + \tau_{nlab}, \quad (6)$$

where τ_{lab} and τ_{nlab} denote the average duration of the labile and non-labile phase, respectively. Fixation durations are the time intervals between successive saccades (excluding saccade execution). In the model, a random walk process delivers the time intervals between two subsequent decisions to initiate a new saccade program and thus influences fixation durations. Figure 2 visualizes the interplay of the different timelines of saccade timing and programming for a string of saccades realized by the model. We highlight two relations: First, due to the autonomous nature of the random walk timing signal, fixation duration does not equate to saccade latency. Second, the time interval between two commands to initiate a saccade program does not translate directly into fixation duration. However, the relationship between the duration covered by the random walk timing signal (see Eq. 1), saccade latency (Eq. 6) and fixation duration is such that increasing the mean duration of the timing signal (t_{sac}) will also lead to prolonged fixation durations (if simulated data are averaged across a sufficient number of realizations).

In the next section, we will further elaborate on how the model architecture obeys the restrictions arising from our knowledge about the timeline of saccade programming and information processing.

On the Timeline of Information Processing

Our working assumption is that fixation durations at least partly reflect moment-to-moment cognitive processes during scene viewing. Mean fixation durations in scene viewing approach 300 ms. It takes roughly 50-60 ms for information about the fixated element to travel from the retina to higher cortical areas (McConkie, 1983). This eye-brain lag and the minimum latency of saccadic eye movements place constraints on the direct-control assumption. Mean estimates for the minimum saccade latency range between 175-250 ms (e.g., Salthouse & Ellis, 1980); a lower boundary of 150-175 ms has been reported for instances where uncertainty about when or where to move the eyes was eliminated (Rayner, Slowiaczek, Clifton, & Bertera, 1983). If we accept the assumption that results from simple visual tasks used in these studies generalize to dynamic tasks like scene viewing, search, or reading, the decision to initiate a new saccade program must be made sufficiently early in the fixation (e.g., Reichle et al., 1998, for reading). In sum, we are left with a very narrow time window during which a sufficient degree of processing of scene elements must be achieved. As an alternative, and this is the approach we adopt here, one could make the case for a somewhat weaker coupling between eye-movement programming and processing of the currently fixated object (e.g., Engbert et al., 2005, for reading). Specifically, the initiation of saccade programs is not intimately tied to a narrow time window within the current fixation. Rather, saccade programs can also be initiated (by the saccade timer) prior to the onset of the current fixation (see below). Strictly

speaking, this proposition is the logical consequence of the hypothesis that the last stimulus information that can influence the programming of the next saccades must occur at least 150 ms (i.e., the minimal saccade latency) before the end of the current fixation. Such a hypothesis implies that fixations shorter than 150 ms reflect instances of saccadic programming beginning before the processing of information from the current fixation (Morrison, 1984; Shioiri, 1993).

Short vs. Long Fixation Durations

Short duration fixations commonly occur in reading (Radach, Heller, & Inhoff, 1999), scanning and visual search (Findlay et al., 2001; McPeek et al., 2000), and scene perception (Henderson & Pierce, 2008; Henderson & Smith, 2009). Generally, the functional role of brief fixations is poorly understood (e.g., Inhoff & Radach, 1998). Our assumption of rhythmic timing and stochasticity in saccade programming suggests that saccades of different latencies and/or fixations across the whole range of durations principally share the same preparatory processes. In empirical research on eye-movement control in visual-cognitive tasks, fixation durations shorter than around 80 ms are frequently discarded from data analysis (see Inhoff & Radach, 1998, for a discussion of cutoff values). The rationale here is that these fixation durations are too short to be controlled by the visual stimulus. Saccades with a latency below 80 ms are commonly classified as anticipatory saccades, i.e., expectancy-driven movements prepared in advanced and executed shortly after target onset. The fastest visually guided saccades, often labeled express saccades, have a latency of around 100 ms (see Fischer & Weber, 1993, for a debate about the properties and significance of express saccades). The subdivision of saccades into anticipatory, express and regular saccades originates from

basic oculomotor studies employing some form of a target-elicited saccade paradigm. In a typical experiment, the participant is asked to make a saccadic orienting movement to a target that appears in some location in the peripheral visual field. An extensively studied version of this type of experiment is the gap paradigm where a temporal gap separates the offset of the central fixation point from the onset of the eccentric target (e.g., Kalesnykas & Hallett, 1987). These paradigms present well-defined tasks with well-defined saccade targets. In comparison, the location of the target location for the next saccade during scene viewing is less well-defined because of the large set of potential candidate targets. In scene viewing, the functional role of anticipatory saccades is thus less clear. Generalization from standard oculomotor aiming tasks to a dynamic situation like scene viewing is further complicated by the fact that in the latter, saccade latency does not necessarily equate to fixation duration because saccade programming may not begin until some time after fixation onset.

With the modeling approach taken here, we propose that fixations across the whole range of durations principally share the same preparatory processes. However, they differ in that fixations with long durations are most likely to reflect instances of direct control whereas those with short durations are less likely to do so (Henderson & Pierce, 2008; Henderson & Smith, 2009). In the present model, variations in fixation durations are generated by a continuous-time random walk. Furthermore, the model takes into account the fact that saccades can be programmed in parallel, i.e., the preparatory processes of two different saccades can overlap in time (Becker & Jürgens, 1979). As a consequence of both assumptions, two cases can be distinguished. First, the program for the next saccade will be started somewhere during the current fixation. These are

instances of saccade programming that are based on the processing of visual-cognitive information from the current fixation. In other words, these are instances of direct control. Second, the model will also generate instances in which the program for the next saccade has already been initiated *before* the start of the current fixation. These are instances of saccade programming in which visual-cognitive information from the current fixation has limited or no influence at all (as the random walk's transition rate is initially adjusted to the processing difficulty of the region fixated at the *previous* fixation and/or saccade programming has already proceeded to the non-labile or execution stage). In Simulation Study 1, we numerically investigate the prevalence of these two cases and how they relate to fixation duration. Simulation Study 3 is an in-depth demonstration of how direct control is realized in the model.

Simulation Studies

Four simulation studies were conducted. The goal of Study 1 was to reproduce the features of a typical empirical fixation duration distribution observed in scene viewing. A further goal of this study was to exemplify some features of the basic model architecture that were of particular relevance for Simulation Studies 3 and 4. Simulation Study 2 explored an important aspect of cognitive control: the global effect of task instruction on the control of fixation durations in scene viewing. Specifically, we compared two tasks that are widely used in the literature, a memorization task and a visual search task. In Simulation Study 3 we focused on the degree to which (and how) fixations in scene viewing are controlled 'on line' by the current visual scene input (Henderson & Pierce, 2008; Henderson & Smith, 2009). Finally, the goal of Simulation Study 4 was to further test the generality of the developed model architecture by reproducing the pattern of

mean fixation durations observed in another paradigm (the mask onset delay paradigm) designed to investigate the accrual of scene information during fixations (Rayner, Smith, Malcolm, & Henderson, 2009).

Simulation Study 1: The Baseline Model

For the most part, fixation duration analyses in the scene perception literature occur at the level of the means. However, changes in mean duration (or the lack thereof) may reflect distinct patterns at the level of underlying distributions. Thus, the primary goal of Simulation Study 1 was to go beyond the mean and reproduce a typical empirical fixation duration distribution observed in scene viewing (Figure 3). The empirical data came from a new scene onset delay experiment replicating Henderson and Pierce (2008; Henderson & Smith, 2009; see Simulation Study 3). Twelve participants each viewed 40 photographs of real-world scenes. During critical fixations the scene only appeared after a manipulated delay. The critical fixations with delays occurred every sixth fixation. Baseline performance was derived from a control condition where the scene was presented without a mask (0-ms delay), and these were the data used for the baseline model simulations.

Based on the basic model architecture outlined above, 12 statistical subjects viewed 40 arbitrary scenes each. Our goal was to determine whether the model could capture the global trends in the experimental data (Figure 3). Therefore, for simplicity model parameters were fit by eye (Table 1). The mean parameter values were informed by findings from basic oculomotor research, which ensured their psychological and/or neurophysiological plausibility. We assume that saccade-programming processes are stochastic in nature. For each realization of the model simulation, parameter values for

the different saccade programming phases (τ_{lab}, τ_{nlab}, and τ_{ex}) were drawn from gamma distributions with means and standard deviations as listed in Table 1 (see Appendix B for details). The mean time interval between two commands to initiate a saccade program, t_{sac}, was set to 250 ms. Parameter N, which determines the variance of the timing signal, was set to 11. The relation between the mean and variance of the timing signal (Eq. 1) yields a random walk change rate of 0.044. In its implementation, the model generated sequences of saccades and fixations, and the simulated fixation duration data were analyzed in the same way as the empirical data. The simulated data captured the overall trend in the empirical data very well (Figure 3). Specifically, they reproduced the characteristic positive skew in fixation duration distributions, with the mode lying below the mean.

Insert Figure 3 about here

Insert Table 1 about here

In Simulation Studies 3 and 4 we will utilize saccade cancellation as a mechanism contributing to prolonged fixation durations. We investigated the groundwork for this concept first using the baseline model. In particular, we explored how the duration of a fixation relates to the timing of the saccade program terminating that particular fixation. In Figure 4a, fixation duration is plotted as a function of the saccade program's start time (t). The start time t was calculated relative to the beginning of the fixation. Positive x

values indicate that the next saccade program was launched during the current fixation (65% of all cases). Negative time values reflect the instances where saccadic programming began before the processing of information from the current fixation (35% of all cases). Data points are plotted with different symbols, which designate the number of saccade program cancellations involved. Black circles represent the default of no saccade program cancellation during the fixation (79%). In 16.5% of all simulated fixations the corresponding saccade program was cancelled once (red squares), and 3.5% were cancelled twice (blue diamonds); all other cases are negligible. One cancellation means that two labile programs were actually involved in generating the fixation duration, but the second more recent one overrode (cancelled) the first one. Figure 4a shows the start time of the labile program that was eventually executed (and thus not cancelled). Saccade cancellation moves the start (time) of the saccade program that eventually terminates the fixation further into that fixation (Figure 4a and c) and prolongs fixation duration (Figure 4a and b). The three distributions in Figure 4c are offset by a value approaching mean saccade latency (Eq. 6, Table 1). As a consequence, the corresponding fixation duration distributions are shifted in a similar manner (Figure 4b). (Note that these are relative frequency distributions, and no saccade cancellation is the default.) Subtracting the saccade program start time t (x value in Figure 4a) from the fixation duration (y value) yields the saccade latency (cf., Eq. 6). For $t = 0$, fixation duration equals saccade latency. Of course, saccade latency distributions do not differ across the cancellation conditions (inset plot in Figure 4a).

In sum, we observed a strong correlation between the launch time of the next saccade program, locked to the onset of the current fixation, and the resulting fixation

duration (r = .95, p < .001). Importantly, short fixation durations originated from instances where the corresponding saccade program was started before the onset of that particular fixation (Figure 4a). Furthermore, the simulations demonstrate that saccade cancellation prolongs fixation durations, which accords with empirical findings (e.g., processing of letter stings: Vergilino-Perez & Beauvillain, 2004; sentence reading: Yang & McConkie, 2001).

Insert Figure 4 about here

Simulation Study 2: Modeling Viewing Task Influences on Fixation Durations in Scene Viewing

The second simulation study was directed toward exploring, at a global level, the role of cognitive factors on saccade timing and programming in scene viewing. This study also addressed a possible concern with the empirical data for the baseline simulation in Study 1, which were taken from the control condition in the scene onset delay paradigm. Therefore, another objective of Study 2 was to model data from a natural viewing task without gaze-contingent manipulations.

In Simulation Study 2, we explored one particular aspect of cognitive control: the influence of viewing task on the control of fixation durations during scene viewing. Such task effects on eye-movement control parameters were first highlighted in two classic eye movement studies (Buswell, 1935; Yarbus, 1967). Though ahead of their time, these studies presented sparse, qualitative data, and viewing task was not manipulated in a

controlled manner. Furthermore, neither of these studies highlighted task differences on fixation durations. In a study by Võ and Henderson (2009), task instructions were manipulated across experiments, and average fixation durations were significantly longer in a memorization experiment than in a visual search experiment (see also Henderson et al., 1999). In another study, viewing task was manipulated within subjects (Castelhano et al., 2009). Viewing task did not affect the average duration of individual fixations. However, longer gaze durations were found on objects in the scenes during memorization than search. Given the findings from these studies, it is currently not clear whether and under what conditions task instructions influence measures of fixation time. Therefore, we set out to reexamine how task instruction influences fixation durations with high statistical power (36 participants, 45 scenes per viewing task). Furthermore, to control for stimulus-based sources of input to the gaze control system, we rotated the images across viewing tasks. If individual fixation durations are sensitive to task instructions, with more encoding time needed for memorization (Henderson et al., 1999; Võ & Henderson, 2009), then fixation durations should be longer for memorization than for search.

Methods

Thirty-six participants (12 males; mean age = 22.2 yrs) were presented 135 unique full-color photographs (800 × 600 pixels) of real-world scenes from a variety of scene categories. Scenes were presented on a 21-inch CRT monitor at a distance of 90 cm for 8 s each. During scene presentation, participants' eye movements were recorded using an SR Research EyeLink 1000 eye tracker. Viewing was binocular, but only the right eye was tracked. The 135 scenes were divided into three blocks of 45 scenes. In each block, participants performed a different viewing task. For the purpose of this paper only two of

the three tasks were analyzed: Search and Memorization.[4] In the Search task participants were presented with a target word prior to scene presentation and pressed a button as soon as the object was located in the scene. Only eye movements up to the button press were included in the analysis. In the Memorization task the participants had to encode the scene to prepare for an old/new recognition test administered at the end of the experiment. Eye movements from the full 8 s presentation period were analyzed during Memorization. Scenes were rotated through task and order across groups of participants.

Behavioral Data

Fixations were excluded if they were preceded by or co-occurred with blinks, were the first or last fixation in a trial, or had durations less than 50 ms or longer than 1200 ms. In the Search task, participants indicated that they had localized the target after an average of 3618 ms (*SD* = 2335 ms). In this period of time, they made on average 12.4 fixations (*SD* = 8.2). In the Memorization task, participants made on average 24.7 fixations (*SD* = 4.2) during the 8-s scene presentation time. A paired two-sample *t*-test revealed that the mean fixation duration during Memory Encoding (*M* = 267 ms, *SD* = 126 ms) was significantly longer than during Search (*M* = 232 ms, *SD* = 96 ms; $t(35) = 12.0$, $p < .001$). An analysis of ordinal fixation number confirmed that this difference was present throughout the entire trial, indicating that it was not an artifact of the longer analysis period in the Memorization task. The influence of viewing task was also evident in the fixation duration distributions (Figure 5a). For the Memorization task, the modal portion of the distribution shifted slightly towards longer fixation durations, and the tail of the distribution increased somewhat.

Insert Figure 5 about here

Simulation Data: Model Adjustments

Modeling the effect of task instructions on fixation duration within the context of CRISP rests on the following basic assumption: The global characteristics of fixation durations in both viewing tasks can be explained by a single model architecture, with task-specific influences realized by different parameter settings. According to the model architecture, the following components contribute toward systematic differences in fixation durations: (1) the mean (t_{sac}) and variance (i.e., number of states N) of the random walk timing signal, (2) the mean duration of the labile saccade program (τ_{lab}), (3) the mean duration of the non-labile saccade program (τ_{nlab}), and (4) the mean duration of saccade execution. The stochastic saccade cancellation mechanism constitutes an additional source of variance in fixation durations. We will now discuss, from a theoretical and empirical point of view, which model parameters may vary or may be held constant across viewing tasks. Put another way, do task instructions influence saccade timing or saccade programming, or both? First, it is reasonable to assume that the mean and variance of the random walk process underlying the initiation of a new saccade program can vary across viewing tasks. Specifically, implementing a more conservative saccade planning strategy (e.g., to ensure sufficient encoding time in a memorization task) would be captured by the model as a higher mean value for the random walk timing signal. Furthermore, task instructions could affect saccade-programming parameters, i.e., saccade latency (Eq. 6). In standard oculomotor aiming tasks, saccade latency can be directly measured as saccadic reaction times (SRTs). SRTs have been shown to be

sensitive to various bottom-up processes arising from changes in sensory information in the visual field (e.g., Kalesnykas & Hallett, 1987). In addition, it has been shown that a top-down process, arising from specific task instructions, can also affect SRTs (Trottier & Pratt, 2005). In natural scene viewing, saccade latency cannot be directly measured, and experimental research needs to establish whether findings from standard oculomotor aiming tasks generalize to natural scene viewing. For the moment, as a working hypothesis we assume that the duration of the labile stage of saccade programming is sensitive to task differences. Finally, the duration of the non-labile stage of saccade programming (also known as saccadic dead time, SDT) is often assumed to remain relatively constant (e.g., Van Loon, Hooge, & Van den Berg, 2002). Therefore, for the simulations reported below the non-labile saccade programming parameter was fixed at 40 ms. Thus, while incorporating some variability in the duration of the non-labile stage of saccade programming (see Table 2), we assume that the distribution is constant across viewing tasks, meaning that the non-labile stage will not contribute to systematic differences in fixation durations. The same is true for saccade execution. Here, the mean parameter value was estimated from the empirical data: The mean saccade duration across viewing tasks was 37 ms, with no significant difference between Memory Encoding and Search ($t(35) = 1.39$, $p = .17$).

Modeling Results

Based on these considerations, numerical simulations were run to determine the best-fitting values for the three free parameters in the CRISP model. The ranges for the parameter values (Table 2) were constrained by findings from basic oculomotor research, ensuring the psychological and neurophysiological plausibility of parameter values. For

parameter estimation, a genetic algorithm (GA) optimization technique was used. The GA minimized a goodness-of-fit measure, which quantified for a given viewing task how much the simulated fixation duration distribution and average fixation duration deviated from the experimentally observed data. The details of the fitting procedure are presented in Appendix B. The obtained minimal values for the deviation measure, Eq. (B3), were $\Delta = 0.30$ (Memorization) and $\Delta = 0.34$ (Search). The corresponding best-fitting parameter values are listed in Table 2. For each viewing task, simulated sequences of fixation durations were obtained from 36 statistical subjects and 45 arbitrary scenes per subject, using the best-fitting values for model parameters.

Insert Table 2 about here

The results of these simulations are summarized in Figure 5. The simulations precisely reproduced the longer mean fixation durations that were observed in Memorization versus Search. They also reproduced the characteristic positive skew in fixation duration distributions, with modal values below the means. Importantly, CRISP captured the global effect of viewing task on fixation durations well: The Search distribution exhibits a somewhat earlier mode and a shallower tail than the distribution for Memory Encoding. On closer inspection, this difference is mainly due to a lower mean value of the random timer and a shorter labile stage of saccade programming in Search than Memorization (Table 2). We offer the following psychological interpretation for the model's prediction of a shorter labile stage of saccade programming in the search task: In search, where observers look for a predetermined target object, knowledge about

the nature of the target and its relationship to the scene can be used to constrain the number of potential target locations competing for selection (Torralba et al., 2006). Assuming that the decision about where to move the eyes next is made towards the end of the labile stage (cf., Becker & Jürgens, 1979; Engbert et al., 2005), simplified target selection would be associated with a shorter labile stage of saccade programming (cf., Rayner et al., 1983). Finally, both the absolute range and the shape of the fixation distribution observed under the memorization instruction in the present study (Figure 5) are in good agreement with the data presented in Study 1 (Figure 3), supporting the validity of these data.

To summarize, Study 2 was concerned with the influence of viewing task on the distribution of fixation durations in scene viewing. We proposed that participants implement different global parameter settings as they examine scenes under different task instructions. We showed that viewing task could be captured by CRISP by allowing task to influence the random walk timer of the saccade generator and the labile stage of saccade programming.

Simulation Study 3: Modeling Mixed Control of Fixation Durations in the Scene Onset Delay Paradigm

Simulation Study 3 concerns a question that is at the heart of current research on fixation durations in scene viewing: Are the decisions about when to move the eyes under direct moment-to-moment control of the current visual scene? We approached this complex question by simulating data from an experiment that selectively manipulated global scene processing difficulty by delaying scene presentation during critical fixations (Henderson & Pierce, 2008; Henderson & Smith, 2009). This paradigm was initially

developed to assess direct control of fixation durations in sentence reading (Morrison, 1984; Rayner & Pollatsek, 1981) and was part of the inspiration for the development of the E-Z Reader model of eye-movement control in reading (Reichle et al., 1998). The general principles of the interaction between visual-cognitive processing and saccade timing represent an open research problem. In the context of this simulation study we propose and test a small set of simple rules for the modulation of saccade timing and saccade programming by visual-cognitive processing. The presentation of this study is organized as follows: We first describe the experiment and provide a short summary of key findings. We then outline how CRISP was modified for the simulation study. This is followed by comparing the obtained simulated data with the empirical data. Finally, we will discuss how the model architecture generates the distinct data pattern observed in the experiment.

Behavioral Data: The Scene Onset Delay Paradigm

Simulation Study 3 was closely patterned after a variant of the scene onset delay paradigm. In the behavioral experiment, 12 participants viewed 40 photographs of real-world scenes in preparation for a later memory test. Every sixth[5] saccade was manipulated in the following way: During the saccade, i.e., periods of visual blindness (Ross, Morrone, Goldberg, & Burr, 2001), the scene was removed from the monitor and replaced by a pattern mask. Therefore, when the eyes landed in the *critical fixation* following this saccade, the scene was no longer visible. For purposes of analysis and simulation, the critical fixation was the first fixation following replacement of the scene with the mask. After a specified amount of time, called the delay, the scene reappeared. The delay was varied with the values 0, 300, 400, 600, 800 ms, or infinite. Delay values

were chosen pseudorandomly for each critical fixation within each scene. In the 0-ms delay control condition, the scene image was replaced with itself so that phenomenally it was continuously present but the computer code generating changes in the other conditions was controlled. The present experiment differed from previously reported experiments (Henderson & Pierce, 2008; Henderson & Smith, 2009) in that it also included an infinite delay condition in which the scene only reappeared when participants moved their eyes to end the critical fixation. Within each scene, delay values were chosen pseudo-randomly for each critical fixation. For details on the stimuli, procedure, and implementation see Henderson and Pierce (2008). If the duration of eye fixations was directly controlled by the current visual input, then fixation durations should systematically scale with the delay. Simply put, the longer the scene is removed from view, the longer the eyes should remain where they are. Figure 6b visualizes the duration of the critical fixation as a function of the realized scene onset delay. The *realized scene onset delay* takes the exact time point of scene disappearance into account; for example, if the display change was completed 8 ms before the beginning of the critical fixation, the resulting scene onset delay for a 400 ms delay trial would amount to 400 – 8 = 392 ms. In Figure 6, the diagonal line represents the 'infinite delay' where fixation duration equals scene onset delay. For the other delay conditions, the empirical data show as two populations (Figure 6b), suggesting that the underlying distributions are bimodal (Figure 9). First, we observed an "early" population of fixations that terminated during scene absence. It appears that these fixation durations line up largely independent of the delay condition. This was confirmed by fitting 2-term Gaussian distributions to the empirical data and performing a regression analysis over the first modes of the fitted distributions

(slope = 0.05, intercept = 243.89, $p > .05$; see Henderson & Pierce, 2008). In addition, we observed a second, "late" population of fixation durations. Here, the fixation duration was longer than the actual delay, meaning that subjects waited for the scene to re-appear. These "waited out" durations linearly increased as the delay increased, which was corroborated by a regression analysis over the second modes of the fitted distributions (slope = 1.06, intercept = 185.71, $p < .01$). This latter population of fixations indicates that fixation durations are, at least partly, influenced by the current visual stimulus. Together, the two populations of fixation durations are indicative of mixed control involving both direct and indirect control of fixation durations. The global pattern of results is consistent with prior results from studies of scene viewing (Henderson & Pierce, 2008; Henderson & Smith, 2009) and reading (Morrison, 1984; Rayner & Pollatsek, 1981).

The two populations of fixation durations appear to be separated by a gap, which is due to saccadic inhibition: The reappearance of the scene creates a motion transient, which inhibits the generation of saccades (Reingold & Stampe, 2002, 2004). Saccadic inhibition has been observed in a broad range of eye-movement tasks including simple saccade programming paradigms (Reingold & Stampe, 2002), reading (e.g., Reingold & Stampe, 2004), visual search (e.g., Reingold & Stampe, 2004; Stampe & Reingold, 2002), and picture viewing (Graupner, Velichkovsky, Pannasch, & Marx, 2007; Pannasch, Dornhoefer, Unema, & Velichkovsky, 2001). In addition, it has been observed as part of the microsaccade-rate signature in various fixation tasks investigating the allocation of attention (see Rolfs, Kliegl, & Engbert, 2008, for a review). In the eye-movement studies reported above, typically a sudden flicker and thus *disruptive*

information is briefly presented at random intervals during fixation. Note that the scenario in our experiment is somewhat different in that the drop in saccadic activity is observed when we reinstate the scene and thus present *useful* information. In sum, the entire pattern of data replicates in essential details the data reported by Henderson and Pierce (2008) and Henderson and Smith (2009).

Insert Figure 6 about here

Simulation Data: Model Adjustments

Our simulation efforts set out to qualitatively reproduce the key behavioral findings. In the implementation, the sequence of events was closely modeled on the experiment (Figure 7). Every sixth saccade, the scene was removed from view for the duration of the delay condition (e.g., 600 ms). Variability in the realized scene onset delay (Figure 6a) was generated based on simulated variations in saccade durations. In order to capture the specific pattern typically shown in SOD experiments, the CRISP baseline model was furnished with a few additional assumptions. They reflect our hypotheses about how the cognitive-oculomotor system might respond to the scene disappearance and subsequent reappearance. Specifically, we assume that (1) current processing demands modulate the random walk's transition rate, and (2) processing difficulty can lead to saccade cancellation. We discuss the details of these assumptions next.

First, we assume that processing difficulties can inhibit and thus modulate saccade timing via modulation of the random walk's transition rate. The idea here is that the

quality of the stimulus (i.e., the amount of information that can be extracted from the stimulus per unit of time) will affect the saccade timer by modulating the statistics of the random walk. Thus, visual-cognitive processing demands automatically adjust saccade timing, although saccade timing itself is not coupled to certain stages of cognitive processing. How plausible is a modulation of the random walk's transition rate? The assumption is consistent with neurophysiological findings. For example, varying the information content of a stimulus (e.g., the signal-to-noise ratio in a motion coherence stimulus) changes the rate of accumulation in neurons in the lateral intraparietal (LIP) area (Roitman & Shadlen, 2002). Accordingly, in our implementation, when the scene is removed from view in the scene onset delay paradigm, the mean random walk transition rate r_1 is considerably reduced. Here, we take an eye-brain lag of 50 ms into account, so 50 ms following scene offset the mean transition rate is reduced from r_1 to r_0 (Table 3). Thus, we slow down the random walk process, which delays the initiation of the next saccade program. Fifty milliseconds after the scene reappears, the rate recovers to its default value r_1 (see Figure 7 for visualization). We assume that information accumulation drops to a low non-zero value because although perceptual uptake ceases during the delay, cognitive processes including conceptual analysis and memory consolidation continue to operate (Potter, 1976). According to the model architecture, no new saccade program would be launched during scene absence if r_0 had a value of 0. The empirical data suggest that this is not the case. Each discrete time step dt was simulated according to Eq. (4), with $w_0 = r_1$ when the scene was visible, and $w_0 = r_0$ if the scene was removed from view. This adds a new free parameter to the model, i.e., the fraction r_0 / r_1 (see Table 3).

Insert Figure 7 about here

We posit saccade cancellation as a second mechanism contributing to prolonged fixation durations in the SOD paradigm. The underlying rationale is that removing the scene from view might interrupt the preparation of eye movements. At the time point of scene disappearance, if there is a labile saccade program active, it is subject to cancellation. Here, noise was added to the cancellation process ensuring that not every cancellation was successful ($p1_{canc} = 0.5$). A similar cancellation mechanism was implemented as response to the scene reappearance to reflect saccadic inhibition (Reingold & Stampe, 2002, 2004). Again, in the simulations the probability of aborting a labile saccade program is smaller than 1 ($p2_{canc} = 0.67$). Note that the impact of such a cancellation mechanism depends on the baseline probability of active labile saccade programs at the time of scene disappearance or reappearance (see Table 4 below for a numerical examination). In sum, the model comprises six parameters related to saccade timing and programming (t_{sac}, N, r_0/ r_1, τ_{lab}, τ_{nlab}, τ_{ex}) as well as two cancellation probabilities, all of which are summarized in Table 3.

Insert Table 3 about here

Modeling Results

Several features of the empirical data pose particular challenges for modeling. First, the simulations need to generate as many prolonged fixations during the onset delay

as those produced by human participants. Second, fixation durations in the experiment are not globally increased, but increase only when the scene is removed from view, i.e., in response to the experimental manipulation. Thus, the empirical data cannot be accounted for by a mere change in mean and variance of the random walk process as global model parameters as was the case in Study 2. Relatedly, the first mode of the bimodal distribution of critical fixation durations shows little variation across the different delay conditions (see Figure 9); instead, the effect of delay condition mainly manifests in the tail of the distributions. Third, while we cannot rule out the possibility that participants anticipated delays, the duration of the delay for a given critical fixation was not predictable. Consequently, the obtained data pattern cannot be explained by a global strategy in which the visuo-oculomotor system adjusts to a constant delay duration (see also Henderson & Pierce, 2008, and Henderson & Smith, 2009, for direct evidence against such a global strategy). Fourth, inclusion of very long delays is a challenge because the model must generate a number of extremely prolonged fixation durations. Finally, the infinite delay condition forms an important boundary condition. In all other experimental conditions, the delay is randomly chosen from a fixed set of delay durations (see above). In the infinite delay condition, however, participants determine the time point of scene reappearance themselves since the scene only reappears when they move their eyes away from the current location to end the critical fixation (theoretically, not making any eye movements would thus lead to an infinite delay). Note that participants are typically not aware of this relationship.

Mean probabilities and durations. Mimicking the experiment, simulated data were obtained from 12 statistical subjects and 40 arbitrary scenes per subject. The

simulated data qualitatively reproduced the two populations of fixation durations observed in the empirical data (Figure 6). Data points below the diagonal represent fixation durations that were terminated before the scene returned to view. Data points above the diagonal reflect fixation durations that waited until the scene returned. There is also a hint of a saccadic inhibition gap, separating the two populations.

We translated this overall pattern into fixation probabilities and durations (Figure 8). Figure 8a shows the probability of "waiting out the mask" as a function of scene onset delay. Across the different delay conditions, CRISP reproduces these probabilities very well. Furthermore, Figure 8b depicts the mean fixation durations across the different delay conditions, separately calculated for the two populations of fixation durations. Again, data points below the diagonal reflect mean fixation durations for all instances where the critical fixation was terminated during the mask. Data points above the diagonal reflect the data points where the duration of the critical fixation was longer than the delay. The figure also shows the empirically observed mean fixation duration observed for the 0-ms delay control condition. This mean duration is well reproduced by CRISP (cf., Simulation Study 1). In addition, CRISP qualitatively reproduces the empirically observed data pattern for the SOD conditions (Figure 8b), but not in a quantitatively exact way - for some of the delay conditions, the observed mean fixation durations are somewhat overestimated or underestimated.

Insert Figure 8 about here

Distributions. Given the clear separation of two populations of fixation durations

in the scatter plot (Figure 6), the underlying frequency distributions must be bimodal. Figure 9 plots the corresponding distributions of fixation durations, separately for each scene onset delay condition. Distributions for the (normal scene viewing) control data are additionally presented. CRISP captured the bimodal fixation duration distributions remarkably well (note that they were generated by a simple processing model rather than resulting from a curve fitting procedure). Because of this qualitative agreement, we did not further optimize the model's goodness-of-fit by implementing an advanced fitting procedure.

Insert Figure 9 about here

Note that in all non-infinite delay conditions, the second peak of the distribution rises about 100 ms following scene reinstatement. The dip in fixation duration distribution is likely due to saccadic inhibition. It has been suggested that saccadic inhibition is due to low-level reflex-like oculomotor processes in response to visual display changes (Reingold & Stampe, 2002, 2004; but see Pannasch et al., 2001). Reingold and Stampe (2002) relate saccadic inhibition effects to inhibitory processes in the superior colliculus. It is further suggested that saccadic inhibition may function to provide the perceptual system with more time to process changes in visual input by delaying the execution of saccades (Reingold & Stampe, 2004). Within the framework of CRISP, we translate this reasoning into a tangible and numerically testable hypothesis. We propose that, in response to the critical event, a large proportion of currently labile saccade programs are cancelled with the effect of prolonging the latency from scene

reinstatement. Numerical simulations showed that implementing this mechanism as part of the model architecture qualitatively accounted for the saccadic inhibition signature observed in response to display changes. CRISP also provides an explanation for unaffected fixation durations that terminate very shortly after the display change: These fixations reflect instances in which saccade programming has already proceeded to the non-labile stage (i.e., the display change occurred after the "point of no return") or saccade execution stage.

The infinite delay condition allows us to determine the extent to which the results of the scene onset delay experiment are contaminated by the saccadic inhibition artifact. In the infinite delay condition, both the initial scene offset and the subsequent scene reappearance take place during a saccade. Due to saccadic suppression (Ross et al., 2001), there are no motion transients to produce saccadic inhibition. Having eliminated saccadic inhibition, we observe a unimodal empirical distribution. Compared to the control data, the distribution is skewed towards longer fixation durations, indicating that fixation durations are under direct control of the current visual input.

Disappearance of the scene. Despite the relative simplicity of CRISP, the manner in which fixation durations are affected by the interplay of random walk timing signals and the different phases of saccade programming is quite complex. However, numerical simulations with the model allow us to "look inside" the generation of fixation durations. As an example, we now take a closer look at saccade programming around 50 ms (reflecting the eye-brain lag) after the scene was removed from view. For this particular point in time, which is close to fixation onset, each simulated realization of a scene onset delay was assigned to one of the five cases listed in Table 4. In 27% of all cases, none of

the simulated saccade programming stages were active when the scene disappeared. (This high percentage is related to the fact that the scene always disappeared during a saccade; see also Figure 4.) In 72% of all cases, a labile saccade program was currently active, which was subject to stochastic cancellation. This saccade cancellation mechanism, together with the reduced random walk transition rate during scene absence, ensured that the model generated the prolonged fixation durations we observed in the SOD paradigm. Furthermore, there were a few cases in which either a non-labile saccade program was active (0.9%) or a saccade was being executed (1 incident) at the critical point in time. In such cases, scene disappearance cannot affect the critical fixation duration at all; typically, these fixation durations will be very short. Finally, since we allow for temporally overlapping programming of saccades, a labile and a non-labile program can occur in parallel, but with the set of parameters used in the simulations (Table 3) this did not happen at the time point of scene disappearance.

Insert Table 4 about here

Discussion of SOD Simulation

We now move from numerical computations to a higher level of abstraction and summarize how the proposed model architecture generates the typical data pattern observed in the scene onset delay paradigm. In particular, any model must reproduce three basic features: (1) one set of fixation durations is largely unaffected by the duration of the delay, (2) another set of prolonged fixation durations exceeds the delay, and (3) there is reduced saccadic frequency following scene reappearance after the delay. First, a

salient feature of the pre-terminated shorter fixation durations is that their mode does not seem to vary systematically as a function of delay duration (Henderson & Pierce, 2008; Henderson & Smith, 2009). Broadly speaking, these pre-terminated fixations are instances of saccade timing and programming that are not affected by the experimental manipulation. Second, how are prolonged fixation durations generated? Reducing the random walk's transition rate during scene absence delays the start of the next saccade program and thus potentially prolongs the duration of the critical fixation. If there were, however, a labile saccade program active at time of scene disappearance, rate reduction alone would not prolong the critical fixation duration. In the model simulations, the duration of the different saccade-programming stages is not affected by the experimental manipulation. An active labile saccade program would thus become non-labile and eventually be executed, terminating the current critical fixation as a "normal fixation". In the SOD paradigm the whole scene is taken away from view, which from the perspective of information processing poses a worst-case scenario. To account for the imposed delay in visual processing, we draw on the flexibility of the visuo-oculomotor system: When programming has only reached an early (labile) stage, the upcoming saccade can still be modified or even cancelled (e.g., Becker & Jürgens, 1979). Therefore, in our implementation, currently active saccade programs are cancelled with probability $p1_{canc}$ (Table 3). This cancellation mechanism, combined with the reduced random walk transition rate will lead to prolonged fixation durations during scene absence.

Simulation Study 3 showed that the described mechanisms were able to generate the empirically observed data pattern. The results were based on a high number of model realizations. At the level of a single model realization, however, the above black-and-

white description of how to obtain the two principal outcomes must be seen in the light of the stochastic features of CRISP. For a substantial proportion of critical fixations, the assignment of the critical fixation to either category (pre-terminated vs. prolonged fixation) is simply a result of the simulated stochasticity in saccade timing and programming processes, while the effect of stochasticity is strongly modulated by delay duration. For shorter delays close to the mean fixation duration in scene viewing, some of the fixations that are longer than the delay are simply due to stochasticity. For the long delay conditions, on the other hand, some of the pre-terminated fixations will actually be instances of direct control.

Third, the two populations of fixation durations are separated by a dip in distributions (Figure 9). The underlying reduction in saccadic activity (i.e., saccadic inhibition) is caused by the motion transient created by the reappearance of the scene. The model accounts for saccadic inhibition by the cancellation of saccade programs.

Simulation Study 4: Modeling Fixation Durations in the Mask Onset Delay Paradigm

The final model simulation revisits the issue of model generalizability. Simply put, generalizability refers to the model's ability to account for more than one effect in one particular task (e.g., Pitt, Myung, & Zhang, 2002). In Simulation Study 4 we developed a strict test of CRISP's generalizability. Based on the core assumptions and parameter set used in the SOD simulations, we extended CRISP to a different experimental paradigm, the *mask onset delay* (MOD) paradigm. This paradigm was first introduced in text reading as the disappearing text paradigm (e.g., Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981; Rayner, Liversedge, White, & Vergilino-Perez, 2003) and was later used with scenes (Rayner et al., 2009; see also van Diepen, Ruelens,

& d'Ydewalle, 1999).

On cursory examination, the SOD and MOD paradigms appear to represent two sides of the same coin. In the SOD paradigm, the scene onset is delayed for a pre-determined amount of time, whereas in the MOD paradigm the onset of a mask following the scene is delayed. However, on closer inspection the two paradigms target different theoretical questions. The SOD paradigm investigates the degree to which individual fixation durations are under direct moment-to-moment control of the current scene. In contrast, the MOD paradigm is directed toward understanding how much time viewers need to encode stimulus properties during eye fixations in scene viewing. In the MOD paradigm, following a manipulated delay in each fixation, the scene is replaced by a mask. The scene then reappears only when the viewer makes an eye movement. The MOD manipulation therefore affects the amount of time viewers can process the scene during a given eye fixation.

The main finding of the MOD experiments conducted by Rayner et al. (2009) was that in order to normally process a scene, viewers needed to see the scene for at least 150 ms during each eye fixation. This was found for both visual search and memorization tasks. Interestingly, the amount of time needed to encode stimulus properties in scene perception is much longer than in reading, as readers need to view the words in the text for only 50 to 60 ms to encode them (e.g., Rayner et al., 1981; Rayner et al., 2003). The MOD simulations in the present paper were based on the memory task included in Experiment 2 of Rayner et al. (2009). We first provide more details on the experimental procedure and results from Rayner et al. (2009), and then present the model simulations.

Behavioral Data

The experiment began with an encoding phase in which each scene was presented for 6 s. Participants were instructed to memorize the scene for a subsequent recognition test. The stimulus material consisted of 60 full-color photographs of real-world scenes from a variety of scene categories. Depending on the mask onset delay condition, at the start of every fixation the scene was visible for 75, 150, 200, or 250 ms before it was masked. Baseline performance was derived from a control condition where no mask appeared. For each participant, each scene was presented in only one mask onset delay condition. After the encoding phase, a memory test was administered (see Rayner et al., 2009, for details).

In the 75 ms mask onset condition (the shortest delay), viewers' accuracy in the recognition memory test was significantly worse than in the control condition; the other mask onset delays did not affect memory performance. Eye-movement measures like fixation duration are indicative of whether normal eye movement behavior had occurred during the encoding phase. Mean fixation durations differed significantly from the control condition in all conditions but the 250 ms mask onset condition (see Fig. 2 in Rayner et al., 2009, data were pooled across tasks).

Simulation Data: Model Adjustments

We now discuss how CRISP can accommodate the pattern of fixation durations observed in the MOD paradigm. In the MOD paradigm, scene presentation time is restricted to a set duration at the beginning of each fixation. This contrasts with the SOD paradigm, where scene presentation duration during critical fixations is delayed by a certain amount of time. In the SOD paradigm, holding fixation until the scene returns to view affords further visual processing of the currently fixated scene. In the SOD model

simulations, two mechanisms were responsible for generating these prolonged fixation durations: In response to the disappearance of the scene, inhibitory signals from the visual-cognitive processing module (1) delayed the start of a new saccade program or (2) canceled saccade programs that were not yet fully specified. Furthermore, some probabilistic saccade cancellation was implemented as a response to the scene reappearance to reflect saccadic inhibition. We propose that only one of these mechanisms is applicable in the MOD paradigm.

In the MOD paradigm, the appearance of the mask signaled the end of scene encoding during the current fixation. The scene only reappeared once the viewer made a saccade to another location. Thus, a reasonable strategy for the viewer would be to move the eyes quickly once the mask appears, to refresh the visual input. However, participants were not informed about the intricacies of the experimental manipulation, and the empirical data were not consistent with this strategy. Instead, mean fixation durations were longer when a mask appeared at a certain point in each fixation. This result suggests that saccadic inhibition plays a major role in prolonging fixation durations in the MOD paradigm. Empirically, both paradigms share the implementation of gaze-contingent display changes during fixations: The reappearance of the scene (SOD) or the disappearance of the scene (MOD) during fixation creates a transient visual event producing saccadic inhibition.

Based on these considerations, in the model simulations we attempted to account for the lengthened fixation durations in the MOD paradigm solely by cancellation of saccade programs to reflect saccadic inhibition. The sequence of events in the model simulations followed the experiment. On a methodological note, Rayner et al.'s (2009)

implementation of the MOD paradigm differs from the SOD paradigm in that every fixation was manipulated. Furthermore, in the SOD paradigm participants faced different scene onset delay conditions in every trial, whereas in the MOD paradigm the mask onset delay did not change within the trial, but only across trials. Thus, in the simulations the mask onset delay was set for a particular trial, and it was either 75, 150, 200, or 250 ms. Every fixation during a given trial was subject to the scene duration. As in the experiment, if a simulated fixation terminated before the deadline set by the mask onset delay condition, it was not considered. This implied that in the conditions with long mask onset delays there were a considerable number of fixations that did not qualify for the manipulation. Finally, a fifth of the trials were control trials with no mask manipulation.

Modeling Results

We consider this simulation study to be a test of the generality of the model and its theoretical principles rather than a data fitting exercise. Therefore, for the simulations we used the same parameters for saccade timing and programming that were used for the SOD data (Table 3, exclusive of r_0 and $p1_{canc}$, because these parameters do not apply). Fifty milliseconds following the simulated disappearance of the scene, labile saccade programs were cancelled with a probability of 0.5. The simulations reproduced the pattern of mean fixation durations observed for the empirical data very well (Figure 10). Fixation durations were longer in the mask onset delay conditions than in the control condition.[6] For the mask onset delay conditions, there was a decrease in fixation duration with each level of mask delay from 75 to 250 ms.

Insert Figure 10 about here

To conclude, the simulations in Study 4 are informative on two counts. On the one hand, Study 4 further demonstrates the model's generalizability. CRISP was able to use parameter values identified from one type of scene viewing paradigm to account for a set of fixation duration data from another scene viewing paradigm. On the other hand, Study 4 demonstrates that the model has predictive power: The simulations predict that mean fixation durations across different mask onset delay conditions should not be different to the control condition if saccadic inhibition is removed.

General Discussion

What we see and understand about the visual world is tightly related to where we place our eyes. Eye movements thus serve as a window into the operation of the attentional system. Eye-movement control during scene viewing can be represented as a series of individual decisions about where and when to move the eyes. Current computational models of attention and gaze control in scene viewing (e.g., Itti & Koch, 2000; Navalpakkam & Itti, 2005; Parkhurst, Law, & Niebur, 2002; Torralba et al., 2006) seek to predict fixation locations ("where"), but they typically ignore fixation durations ("when"). However, conclusions about the distribution of attention over time in scenes can differ markedly when fixation duration is taken into account (Henderson, 2003, 2007; see also Henderson & Smith, 2009). With the present work, we propose a model that accounts for saccade timing and programming and thus for variations in fixation durations during scene viewing. Because our CRISP model incorporates a saccade-programming module, it explicitly acknowledges the restrictions that arise from the operation of the oculomotor system, a contrast with other current models. Although in

this initial model we consider saccade timing without taking fixation positions into account, our longer-range goal is to incorporate a saccade-generation module of the sort proposed here within a more complete computational model of scene viewing. We believe that this incremental approach is reasonable given the relative independence of "when" and "where" decisions in eye-movement control (see Findlay & Walker, 1999).

The model architecture for CRISP can be summarized with three main principles. First, timing signals for saccades are generated by random walks (cf., Ratcliff & Rouder, 1998; Reddi, Asrress, & Carpenter, 2003; Roitman & Shadlen, 2002). Second, difficulties at the level of visual and cognitive processing inhibit (i.e., delay) saccade initiation, essentially leading to longer fixation durations. Inhibition can take two forms: (1) current processing demands immediately modulate the random walk's transition rate, and (2) processing difficulties can lead to saccade cancellation (cf., Vergilino-Perez, Collins, & Doré-Mazars, 2004; Yang & McConkie, 2001). Third, saccade programming is completed in two stages. During the first labile stage the program can still be modified. The program then enters a non-labile stage, which inevitably leads to saccade execution. This distinction is motivated by basic oculomotor research indicating that more than one saccade can be programmed at a time (e.g., Becker & Jürgens, 1979). We conducted four simulation studies to test a model based on these three principles.

The goal of Simulation Study 1 was to perform a global test of the basic model architecture. The results showed that the basic principles and stochastic elements in the model were capable of accounting for the range of fixation durations observed in scene viewing. Simulation Study 2 explored the global effect of task instruction on the control of fixation durations in scene viewing. Specifically, we compared two tasks that are

widely used in the literature, memorization and visual search. For the memorization task, the modal portion of the fixation duration distribution was found to be shifted towards longer fixations, and the tail of the distribution was somewhat increased. The corresponding pattern of average individual fixation durations supports prior research comparing tasks across different groups of participants (Henderson et al., 1999; Võ & Henderson, 2009; but see Castelhano et al., 2009, comparing tasks within participants). The general approach to modeling these data was to realize global task-specific influences by different model parameter settings. The decision on which model parameters to vary or keep fixed across tasks was guided by findings from empirical research and theoretical considerations. The simulated data were in close agreement with the empirical data. These results also present a first demonstration of CRISP's generalizability.

Simulation Study 3 provided further insight into the control of fixations, particularly whether and how fixations in scene viewing are controlled directly by the current visual scene input. Here, the CRISP model was validated by data from a scene onset delay experiment. During selected saccades (when vision is suppressed), the entire scene was replaced by a mask. During the subsequent fixation, the reappearance of the scene was delayed for a variable amount of time. In the participant data, a first population of pre-terminated fixations was largely unaffected by this experimental manipulation. For a second population of fixations, however, fixation duration systematically increased with the delay, suggesting direct control of fixation durations. The model was able to generate this same pattern of data. According to the model simulations, whether or not delaying scene information influences the critical fixation duration depends on how far along

saccade planning has proceeded (cf., Morrison, 1984). Sometimes it will be too late to allow any influence. The saccade will be irrevocably executed, resulting in a comparatively short fixation that ends during the mask. At other times inhibitory signals from the visual-cognitive processing module will delay the start of a new saccade program, eventually prolonging fixation durations. In addition, the participant data reflected saccadic inhibition due to visible changes in the display (Reingold & Stampe, 2002, 2004). Our simulations were able to capture this phenomenon via a saccade cancellation mechanism.

In Simulation Study 4, the numerical model used to account for the SOD data was applied to data from a mask onset delay experiment (Rayner et al., 2009). This experimental paradigm is typically used to study the amount of time viewers need to encode stimulus properties during eye fixations. In the Rayner et al. (2009) experiment that we modeled, the scene was present at the beginning of each fixation but was replaced by a mask after a set delay. The scene then reappeared once the viewer made an eye movement. Thus, the MOD manipulation affected the amount of time viewers could process the scene during a given eye fixation. Rayner et al. (2009) showed that viewers needed to see the scene for at least 150 ms during each eye fixation to produce normal search and memory performance. More critically for our purposes, fixation durations increased as mask onset delay (scene presentation time) decreased. Our model captured the fixation duration data from the memorization task included in Rayner et al. (2009) and replicated the distinct pattern of mean fixation durations observed in the empirical data. According to the assumptions that were implemented in CRISP, the prolonged fixation durations in the mask onset delay conditions were mainly due to saccadic

inhibition associated with the onset of the mask, thereby supporting speculations by Rayner et al. (2009).

Eye-movement control theories can be contrasted as direct (e.g., Reichle et al., 1998) vs. indirect (e.g., Hooge & Erkelens, 1998) theories according to the nature of the mechanism proposed to control fixation durations. CRISP offers an interesting alternative to this theoretical classification. In the model, the visual-cognitive processing module *directly* controls both the timer (via modulation of the random walk's transition rate) and the saccade-programming module (via saccade cancellation). The control mechanisms are currently implemented as inhibitory signals. However, due to conceptualization of saccade timing and programming in the model, the model output will always mimic a *mixed* control of fixation durations.

Note that the key signature of the stimulus onset delay paradigm has been observed across different domains like reading (Morrison, 1984; Rayner & Pollatsek, 1981), scene perception (Henderson & Pierce, 2008; Henderson & Smith, 2009) and visual search (Vaughan, 1982; Vaughan & Graefe, 1977). In its current implementation CRISP does not make assumptions that are specific to scene perception. Our modeling efforts are guided by the principle of model generalizability (e.g., Pitt et al., 2002), i.e., how we control eye movements in scene perception should principally be in agreement with the main control principles guiding the eyes during other tasks like visual search or reading (cf., Engbert et al., 2005). We acknowledge that certain sub-processes are task-specific, and future extensions of the model proposed here must take these into account (see below). Yet, our principle modeling approach clearly aims at a constructive convergence of models across scene viewing and reading (Henderson & Smith, 2009; see

also Nuthmann & Engbert, 2009).

Relation to Other Models

In this section we address some conceptual similarities and differences between CRISP and other models. Our assumptions about saccade programming are consistent with evidence concerning basic oculomotor control (e.g., Becker & Jürgens, 1979), and the notion of two-stage saccade programming has previously been adopted by two fully implemented models of eye-movement control in reading: SWIFT (Engbert et al., 2002; Engbert et al., 2005) and E-Z Reader (Reichle et al., 1998; Reichle et al., 2003). In addition, we assume autonomous saccade timing, which is modulated by processing difficulty. This core assumption is inspired by the SWIFT model of saccade generation (Engbert et al., 2002; Engbert et al., 2005).[7] However, CRISP and SWIFT differ in their conceptualization and implementation of the timer. In SWIFT, autonomous timing signals are drawn from a gamma distribution, but the sampled time interval can be prolonged by a high value of the activation of the currently fixated word. This mechanism is referred to as foveal inhibition of the autonomous timer (Engbert et al., 2005). In contrast, in the present model timing signals are implemented as random walks. There is an obvious advantage of such a conceptualization: The random walk creates a trajectory over time, which can be modulated at any point by the quality of the stimulus. In contrast to other gaze-control models, the present model thus proposes *continuous* crosstalk between saccade preparation and visual-cognitive processing. This crosstalk provides a powerful mechanism for implementing direct control. In addition, we introduce the idea that processing difficulties can lead to saccade cancellation (cf., Vergilino-Perez et al., 2004; Yang & McConkie, 2001).

Our notion of a random walk timing process bears some similarities to both random walk and diffusion models in the manual reaction time domain as well as to a group of accumulator models developed to explain neurophysiological and behavioral data related to the initiation of saccades (see Smith & Ratcliff, 2004, for a review). Like simple perceptual two-choice decisions in manual reaction time studies, saccade generation can be modeled as a process of accumulating evidence up to some response threshold. Saccades are triggered once the signal generated in response to a stimulus reaches the threshold. Models of this type include three sources to explain the stochastic variability observed in saccadic reaction times: (1) variable baseline (e.g., Trappenberg, Dorris, Munoz, & Klein, 2001), (2) variable threshold (e.g., Nazir & Jacobs, 1991), and (3) variable rate of rise from baseline to threshold (e.g., Carpenter & Williams, 1995; see also Ratcliff, 1978). The timing mechanism implemented in CRISP falls into the last category (for neurophysiological evidence in favor of such a fixed-threshold, variable-rate architecture see Hanes & Schall, 1996).

However, scene viewing is a markedly different task than the type of task typically used in basic oculomotor or neurophysiological research on saccade initiation. These differences in task translate into differences in the type of response elicited when reaching threshold. In scene viewing, sequences of saccades are made spontaneously while scanning a static or dynamic image. In basic oculomotor or neurophysiological studies, subjects (humans or monkeys) make saccadic eye movements to indicate their decisions about visual stimuli. Thus, the dependent variable (as in the LATER model, e.g. Carpenter & Williams, 1995; Reddi & Carpenter, 2000) is the time between the appearance of a suddenly presented visual stimulus and saccade initiation. These

approaches therefore consider saccade latencies and not fixation durations. The variability in saccade latencies originates from variability in the time taken for some expectation (as in the LATER model) or neural activity (e.g. Schall, 2004) to reach the threshold. Once the threshold is reached a saccade is executed.

This situation starkly contrasts with the role of the threshold in the CRISP model. Reaching the threshold in CRISP activates a new labile saccade program, not saccade execution. This creates two separate though strongly interacting time lines that combine to produce fixation durations: random walk timing signals and saccade programming involving labile and non-labile stages (Figure 2). At the current fixation location, if the random walk timing signal reaches a certain threshold, a new labile saccade program is started. Inhibitory signals from the visual-cognitive processing module can (1) delay the start of a new saccade program or (2) cancel saccade programs that are not yet fully specified. In sum, such a set-up acknowledges a certain sensitivity of the visuo-oculomotor system to visual-cognitive processing demands. For scene viewing, where saccades are usually not evoked by sudden visual appearances, this seems psychologically more plausible than relating the positive response boundary to saccade execution.

Directions for Future Research

The SOD paradigm allows us to determine whether and how fixation durations are prolonged when the scene input is delayed. In such an experiment, the input is manipulated in an all-or-nothing manner (scene on or off). To go beyond this dichotomy, we are currently developing methods for *parametrically* manipulating the quality of scene input and examining the resulting graded effects on fixation durations. These

manipulations come closer to mimicking the slow-down in processing that might result from more difficult visual or semantic processing of the scene. Such variation would also exploit the full potential of the random walk saccade timer implemented in CRISP: The transition rate of the random walk can locally adjust to the visual-cognitive complexity of the currently fixated scene region.

The majority of fixations during scene perception are centered on or very close to objects rather than on empty space or background (Buswell, 1935; Yarbus, 1967). The challenge for experimental and computational research is to obtain an aggregated measure reflecting the processing difficulty of an object in a scene or a scene region similar to measures like word frequency and word predictability in reading. Consideration of this issue also invites exploration of the interface between eye-movement control in scene viewing and object recognition. Here, a possible point of departure is the descriptive model of object fixation times in scene viewing sketched out by De Graef (2007), suggesting that the visual system monitors the rate of activation buildup in an object lexicon.

In addition, a realistic model of scene viewing must take the constraints arising from acuity limitations of the visual system into account. This is necessary to dissociate foveal from parafoveal and peripheral influences on fixation duration. Fine detail discriminations require foveal processing (cf., Henderson, McClure, Pierce, & Schrock, 1997). However, extrafoveal information is used to obtain global image characteristics, and to guide saccades (e.g., van Diepen & d'Ydewalle, 2003). Examples for implementing visual acuity limitations include the visual map in the *ideal searcher* (Najemnik & Geisler, 2005), and the gaze-contingent foveation filter which was added to

a variant of the saliency model of visual attention (Itti, 2006).

A closer look at current models of attention and gaze control in scene viewing reveals a certain asymmetry. Current models seek to predict fixation locations and ignore timing (see Rayner, 2009a; 2009b, for discussion). The novel contribution of the present work is to provide a model that accounts for saccade timing and programming and thus for variations in fixation durations in scene viewing.

References

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research, 46*(18), 2824-2833.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*, 364-390.

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research, 19*, 1967-1983.

Buswell, G. T. (1935). *How people look at pictures*. Chicago: University of Chicago Press.

Camalier, C. R., Gotler, A., Murthy, A., Thompson, K. G., Logan, G. D., Palmeri, T. J., et al. (2007). Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Research, 47*(16), 2187-2211.

Carpenter, R. H. S., & Williams, M. L. L. (1995). Neural computation of log likelihood in the control of saccadic eye movements. *Nature, 377*, 59-62.

Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology, 62*(1), 1-14.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9(3):6*, 1-15.

De Graef, P. (2007). Transsaccadic recognition in scene exploration. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 165-191). Oxford: Elsevier.

De Graef, P., Christiaens, D., & D'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52*(4), 317-329.

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature, 403*(6767), 335-338.

Engbert, R., & Kliegl, R. (2001). Mathematical models of eye movements in reading: a possible role for autonomous saccades. *Biological Cybernetics, 85*, 77-87.

Engbert, R., & Kliegl, R. (2003). Noise-enhanced performance in reading. *Neurocomputing, 50*, 473-478.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*(5), 621-636.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*(4), 777-813.

Findlay, J. M., Brown, V., & Gilchrist, I. D. (2001). Saccade target selection in visual search: the effect of information from the previous fixation. *Vision Research, 41*(1), 87-95.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences, 22*(4), 661-721.

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences, 16*(3), 553-567.

Fricke, T., & Schnakenberg, J. (1991). Monte-Carlo simulation of an inhomogeneous reaction-diffusion system in the biophysics of receptor-cells. *Zeitschrift Für Physik B-Condensed Matter, 83*(2), 277-284.

Friedman, A. (1979). Framing pictures - The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*(3), 316-355.

Gardiner, C. W. (2004). *Handbook of stochastic methods* (3rd ed.). Berlin, Heidelberg, New York: Springer.

Gillespie, D. T. (1976). General method for numerically simulating stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics, 22*(4), 403-434.

Gillespie, D. T. (1978). Monte Carlo simulation of random walks with residence time dependent transition probability rates. *Journal of Computational Physics, 28*(3), 395-407.

Graupner, S. T., Velichkovsky, B. M., Pannasch, S., & Marx, J. (2007). Surprise, surprise: Two distinct components in the visually evoked distractor effect. *Psychophysiology, 44*(2), 251-261.

Hanes, D. P., Patterson, W. F., & Schall, J. D. (1998). Role of frontal eye fields in countermanding saccades: Visual, movement, and fixation activity. *Journal of Neurophysiology, 79*(2), 817-834.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274*(5286), 427-430.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11), 498-504.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16*(4), 219-222.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537-562). Oxford: Elsevier.

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 417-429.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: an overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269-283). Oxford: Elsevier.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review, 16*(5), 850-856.

Henderson, J. M., McClure, K. K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception & Psychophysics, 59*(3), 323-346.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15*(3), 566-573.

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Evidence from a scene onset delay paradigm. *Visual Cognition, 17*(6), 1055-1082.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210-228.

Holland, J. H. (1992). *Adaptation in natural and artificial systems*. Cambridge (Mass.): MIT Press.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review, 8*(4), 761-768.

Hooge, I. T. C., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research, 38*(9), 1295-1302.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29-53). Oxford: Elsevier.

Itti, L. (2006). Quantitative modelling of perceptual salience at human eye position. *Visual Cognition, 14*(4/5/6/7/8), 959-984.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194-203.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254-1259.

Kalesnykas, R. P., & Hallett, P. E. (1987). The differentiation of visually guided and anticipatory saccades in gap and overlap paradigms. *Experimental Brain Research, 68*(1), 115-121.

Loftus, G. R. (1985). Picture perception: Effects of luminance on available information and information-extraction rate. *Journal of Experimental Psychology: General, 114*(3), 342-356.

Loftus, G. R., Kaufman, L., Nishimoto, T., & Ruthruff, E. (1992). Effects of visual degradation on eye-fixation durations, perceptual processing, and long-term visual memory. In K. Rayner (Ed.), *Eye movements and visual cognition: scene perception and reading* (pp. 203-226). New York: Springer.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*(4), 565-572.

Ludwig, C. J. H., Mildinhall, J. W., & Gilchrist, I. D. (2007). A population coding account for systematic variation in saccadic dead time. *Journal of Neurophysiology, 97*(1), 795-805.

Mannan, S., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9*(3), 363-386.

McAuley, J. H., Rothwell, J. C., & Marsden, C. D. (1999). Human anticipatory eye movements may reflect rhythmic central nervous activity. *Neuroscience, 94*(2), 339-350.

McConkie, G. W. (1983). Eye movements and perception during reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 65-96). New York: Academic Press.

McPeek, R. M., Skavenski, A. A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision Research, 40*(18), 2499-2516.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 667-682.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*(7031), 387-391.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*(2), 205-231.

Nazir, T. A., & Jacobs, A. M. (1991). The effects of target discriminability and retinal eccentricity on saccade latencies - An analysis in terms of variable-criterion theory. *Psychological Research-Psychologische Forschung, 53*(4), 281-289.

Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: An analysis based on the SWIFT model of eye-movement control. *Vision Research, 49*, 322-336.

Pannasch, S., Dornhoefer, S. M., Unema, P. J. A., & Velichkovsky, B. M. (2001). The omnipresent prolongation of visual fixations: saccades are inhibited by changes in situation and in subject's activity. *Vision Research, 41*(25-26), 3345-3351.

Paré, M., & Hanes, D. P. (2003). Controlled movement processing: Superior colliculus activity associated with countermanded saccades. *Journal of Neuroscience, 23*(16), 6480-6489.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107-123.

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16*(2), 125-154.

Pitt, M. A., Myung, I. J., & Zhang, S. B. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*(3), 472-491.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology-Human Learning and Memory, 2*(5), 509-522.

Radach, R., Heller, D., & Inhoff, A. W. (1999). Occurrence and function of very short fixation durations in reading. In W. Becker, H. Deubel & T. Mergner (Eds.), *Current Oculomotor Research: Physiological and Psychological Aspects* (pp. 321-331). New York: Plenum.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59-108.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347-356.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K. (2009a). Eye movements in reading: Models and data. *Journal of Eye Movement Research, 2*(5), 1-10.

Rayner, K. (2009b). The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506.

Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance, 7*(1), 167-179.

Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive control of eye movements. *Psychological Science, 14*, 385-389.

Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology, 33A*, 351-373.

Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 912-922.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science, 20*(1), 6-10.

Reddi, B. A. J., Asrress, K. N., & Carpenter, R. H. S. (2003). Accuracy, information, and response time in a saccadic decision task. *Journal of Neurophysiology, 90*(5), 3538-3546.

Reddi, B. A. J., & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience, 3*(8), 827-830.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125-157.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445-526.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network-Computation in Neural Systems, 10*(4), 341-350.

Reingold, E. M., & Stampe, D. M. (2002). Saccadic inhibition in voluntary and reflexive saccades. *Journal of Cognitive Neuroscience, 14*(3), 371-388.

Reingold, E. M., & Stampe, D. M. (2004). Saccadic inhibition in reading. *Journal of Experimental Psychology: Human Perception and Performance, 30*(1), 194-211.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience, 22*(21), 9475-9489.

Rolfs, M., Kliegl, R., & Engbert, R. (2008). Toward a model of microsaccade generation: The case of microsaccadic inhibition. *Journal of Vision, 8(11):5*, 1-23.

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences, 24*(2), 113-121.

Salthouse, T. A., & Ellis, C. L. (1980). Determinants of eye-fixation duration. *American Journal of Psychology, 93*(2), 207-234.

Schall, J. D. (2004). On building a bridge between brain and behavior. *Annual Review of Psychology, 55*, 23-50.

Shioiri, S. (1993). Postsaccadic processing of the retinal image during picture scanning. *Perception & Psychophysics, 53*(3), 305-314.

Sivanandam, S. N., & Deepa, S. N. (2007). *Introduction to genetic algorithms*. Berlin Heidelberg New York: Springer.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27*(3), 161-168.

Stampe, D. M., & Reingold, E. M. (2002). Influence of stimulus characteristics on the latency of saccadic inhibition. In J. Hyönä, D. P. Munoz, W. Heide & R. Radach (Eds.), *The brain's eye: Neurobiological and clinical aspects of oculomotor research* (pp. 73-87). Oxford: Elsevier.

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research, 46*(12), 1857-1862.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. *Psychological Review, 113*, 766-786.

Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience, 13*(2), 256-271.

Trottier, L., & Pratt, J. (2005). Visual processing of targets can reduce saccadic latencies. *Vision Research, 45*(11), 1349-1354.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology, 59*(11), 1931-1949.

van Diepen, P. M. J., & d'Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition, 10*(1), 79-100.

van Diepen, P. M. J., Ruelens, L., & d'Ydewalle, C. (1999). Brief foveal masking during scene perception. *Acta Psychologica, 101*(1), 91-103.

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: Moving masks and moving windows. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 337-355). Oxford: Elsevier.

Van Loon, E. M., Hooge, I. T. C., & Van den Berg, A. V. (2002). The timing of sequences of saccades in visual search. *Proceedings of the Royal Society of London Series B-Biological Sciences, 269*(1500), 1571-1579.

Vaughan, J. (1982). Control of fixation duration in visual search and memory search: Another look. *Journal of Experimental Psychology: Human Perception and Performance, 8*(5), 709-723.

Vaughan, J., & Graefe, T. M. (1977). Delay of stimulus presentation after saccade in visual search. *Perception & Psychophysics, 22*(2), 201-205.

Vergilino-Perez, D., & Beauvillain, C. (2004). The ability of the saccadic system to change motor plans in scanning letter strings. *Psychonomic Bulletin & Review, 11(2)*, 332-337.

Vergilino-Perez, D., Collins, T., & Doré-Mazars, K. (2004). Decision and metrics of refixations in reading isolated words. *Vision Research, 44*(17), 2009-2017.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision, 9(3):24*, 1-15.

Walker, R., & McSorley, E. (2006). The parallel programming of voluntary and reflexive saccades. *Vision Research, 46*(13), 2082-2093.

Yang, S.-N. (2006). An oculomotor-based model of eye movements in reading: The competition/interaction model. *Cognitive Systems Research, 7*(1), 56-69.

Yang, S.-N., & McConkie, G. W. (2001). Eye movements during reading: a theory of saccade initiation times. *Vision Research, 41*(25-26), 3567-3585.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.

Appendix A

Details on the Random Walk Process

The elementary transitions of the random walk process are characterized by exponential waiting times, Eq. (2). For the exponential probability distribution there is only a single free parameter, which is equal to the mean and standard deviation of the distribution, $\mu_0 = \sigma_0$. For N statistically independent transitions, means and variances add up to the overall mean μ_1 and standard deviation σ_1 of the first-passage time for reaching the threshold N, i.e.,

(mean value) $\mu_1 = N\,\mu_0$

(standard deviation) $\sigma_1^2 = N\sigma_0^2 \quad \Rightarrow \quad \sigma_1 = \sqrt{N}\sigma_0$

Therefore, the relation between standard deviation and mean of the saccade timer with N states is

$$\frac{\sigma_1}{\mu_1} = \frac{\sqrt{N}\,\sigma_0}{N\,\mu_0} = \frac{1}{\sqrt{N}}$$

Appendix B

Some Details of the Simulations

The programs for the model simulations were written in MATLAB (The MathWorks). Simulated sequences of fixation durations were output to an external file. Simulated data were then analyzed in the same way as the empirical data. Variations in timing signals were modeled by random walks (see main text and Appendix A). For simulating the duration of saccade programming stages, gamma distributions were used because response latencies are often approximately distributed as gamma distributions (cf. Reichle et al., 1998). Component gamma distributions were constructed by convolving nine exponential distributions, which means that the histograms of the gamma distributions were unimodal and approximately normal, but with a positive skew.

Details on Fitting Model Parameters in Simulation Study 2

The adequacy of parameter settings was assessed by calculating a goodness-of-fit measure characterizing fixation duration distributions and average fixation durations. What is measured is how much a model's predictions deviate from the experimentally observed data. The root mean squared error was calculated as a deviation measure for fixation duration distributions. For this purpose, for each viewing task fixation duration distributions were calculated from 48 bins with bin centers ranging from 12.5 to 1187.5 ms in steps of 25 ms. In Eq. (B1), the corresponding values obtained from model simulations are denoted by D_k, with the subscript indicating the bin (k = 1, 2, 3, …, 48). The experimentally observed values are denoted by $\overline{D}_k$. The deviation between simulated and experimentally observed data is calculated as

$$\Delta_D = \sum_{k=1}^{48} \left(\frac{D_k - \overline{D_k}}{\overline{D_k} + 1} \right)^2 \quad \text{(B1)}$$

The performance of the model was further evaluated by the mean squared normalized error for average fixation duration in a given viewing task (cf., Reichle et al., 1998). The difference between the predicted (M) and observed ($\overline{M}$) values was squared, and this was then divided by the standard deviation of the experimentally observed values (Eq. B2).

$$\Delta_M = \frac{(M - \overline{M})^2}{\sigma(\overline{M})} \quad \text{(B2)}$$

These two measures were combined to a single deviation measure,

$$\Delta = 1000 * \Delta_D + \Delta_M . \quad \text{(B3)}$$

The optimization problem can be defined as minimizing this deviation, i.e., the parameter combination that provides the best fit (i.e. smallest deviation) is favored. To solve the optimization problem, a genetic algorithm (GA) approach was used (Holland, 1992; Sivanandam & Deepa, 2007). The genetic algorithm repeatedly modifies a population of individual solutions, based on natural selection. During each iteration, individuals from the current population are selected to be parents for the next generation. Three main types of rules are used to create the next generation. First, selection rules select the individuals, called parents, that contribute to the population for the next generation. Second, crossover rules combine two parents to produce children for the next generation. Third, mutation rules apply random changes to individual parents to produce children. Over successive generations, the population evolves toward an optimal solution.

Parameter estimation was implemented in MATLAB, using the Optimization Toolbox, and the Genetic Algorithm and Direct Search Toolbox. A population of 100

combinations of parameter values was iterated over 50 generations. The crossover fraction was set to 0.7 (default 0.8). It denotes the fraction of individuals in the next generation, other than elite children, that are created by crossover. For all other MATLAB GA options, the defaults were used. For a given viewing task, several runs of the GA were used to test the reliability of the estimates for the model parameters. The best-fitting parameter values given in Table 2 represent mean values over 10 runs of the GA per viewing task. Using the model parameter values from Table 2, separate simulations were performed to produce the data shown in Figure 5. MATLAB's Parallel Processing Toolbox was used to solve the estimation problems on a multicore computer, i.e., a 2 x 2.8 GHz Quad-Core Mac Pro. A single run for the parameter estimation took approximately 7 hours.

Acknowledgements

This research was supported by a grant from the Economic and Social Research Council (ESRC) of the UK to JMH (RES-062-23-1092). Portions of this research were presented at the annual meeting of the Vision Science Society (Naples, May 2008) and the annual meeting of the Psychonomic Society (Chicago, November 2008). We thank Erik D. Reichle, Peter De Graef, Geoffrey Underwood, and two anonymous reviewers for many insightful comments on a draft of this article.

Footnotes

[1] The sum of all fixations from first entry to first exit in a region.

[2] In the current implementation of the SWIFT model, foveal inhibition modulates fixation durations with a temporal delay (Engbert et al., 2005).

[3] In this important feature the CRISP model differs from the LATER model of saccadic reaction times (Carpenter & Williams, 1995).

[4] In the third task, participants were asked to evaluate how much they liked each scene. Average fixation duration in this condition did not significantly differ from the mean fixation duration in the memorization task.

[5] This is different from the previous studies where every 10^{th} saccade was manipulated (Henderson & Pierce, 2008; Henderson & Smith, 2009).

[6] Some very long fixations, longer than 1200 ms, were excluded from the empirical control data.

[7] Autonomous triggering of saccades is also proposed by the C/I model (Yang, 2006; Yang & McConkie, 2001), which is an extension of a descriptive model proposed by Findlay & Walker (1999).

Table 1

Model Parameters for Simulations with the Baseline Model

Parameter	Symbol	Mean	Standard deviation	Comment
Saccade timing (random walk)				
default random walk transition rate	$r_1 = \frac{N}{t_{sac}}$	$t_{sac} = 250$	$N = 11$	Eq. 1
Saccade programming				
labile stage (ms)	τ_{lab}	180	1/4*mean	
non-labile stage (ms)	τ_{nlab}	40	1/4*mean	
saccade execution (ms)	τ_{ex}	40	1/3*mean	

Table 2

Model Parameters for Memorization versus Visual Search

Parameter	Symbol	Range	Memorization	Visual Search
Saccade timing (random walk)				
mean	t_{sac}	150-300	253	240
variance	N	5-20	18	17
Saccade programming				
labile stage (ms)	τ_{lab}	80-250	183	156

Note. The non-labile saccade programming parameter (τ_{nlab}) was fixed at 40 ms. The mean duration of saccade execution was estimated from the empirical data (τ_{ex} = 37 ms). To simulate variations in saccade programming parameters (τ_{lab}, τ_{nlab}, τ_{ex}), a gamma distribution with a relation between standard deviation and mean of 1/3 was used.

Table 3

Model Parameters for Modeling of Fixation Durations in the Scene Onset Delay (SOD) Paradigm

Parameter	Symbol	Mean	Standard deviation	Comment
Saccade timing (random walk)				
default random walk transition rate	$r_1 = \frac{N}{t_{sac}}$	$t_{sac} = 250$	$N = 17$	Eq. 1
random walk transition rate during scene absence	r_0	$r_0 = 0.30 * r_1$		SOD
Saccade programming				
labile stage (ms)	τ_{lab}	180	$\frac{1}{3} * mean$	
non-labile stage (ms)	τ_{nlab}	40	$\frac{1}{3} * mean$	
saccade execution (ms)	τ_{ex}	40	$\frac{1}{3} * mean$	
probability of saccade cancellation at scene disappearance	$p1_{canc}$	0.5		SOD
probability of saccade cancellation at scene reappearance	$p2_{canc}$	0.67		SOD

Table 4

Scene Onset Delay Simulation Study: Saccade Programming Scenario at the Time Point of Scene Disappearance

Verbal description	Labile stage	Non-labile stage	Saccade execution stage	Number of cases	%
None of the saccade programming stages active	0	0	0	1429	27.1
Active labile saccade program	1	0	0	3803	72.0
Active non-labile saccade program	0	1	0	46	0.9
Saccade in execution	0	0	1	1	0.0
A labile and a non-labile program in parallel	1	1	0	0	0.0

Note. At the time point of scene disappearance, each simulated realization of a scene onset delay was assigned to one of five cases that can occur. For each case, the combined activity of the different saccade programming stages is binary coded (1 active, 0 not active), while the left-most column provides a verbal description.

Figure Captions

Figure 1. Model overview.

Figure 2. Temporal scheme of saccade timing and saccade programming. The random walk timing signal accumulates towards a positive response boundary (i.e., a threshold). Once the threshold is reached, a response, that is a new saccade program, is initiated. The saccade program enters a labile stage (τ_{lab}), which signals the engagement of the oculomotor system. At the end of the labile stage, a point of no return is reached. During the following non-labile stage (τ_{nlab}), the saccade can no longer be cancelled. Finally, the saccade is executed (τ_{ex}). Fixation durations are the time intervals between successive saccades.

Figure 3. Simulations with the baseline model. Fixation duration distributions for simulated (solid line) as compared to empirical (broken line) data.

Figure 4. Simulations with the baseline model. (a) Each data point represents a simulated fixation duration. The scatter plot displays fixation duration (y axis) as a function of the start time of the saccade program that terminated the fixation. The start time is calculated relative to fixation onset. Color and symbol type of the data points are related to saccade program cancellation (see text for details). Panels (b) and (c) show the corresponding relative frequency distributions while the inset plot in panel (a) displays saccade latency distributions.

Figure 5. Viewing task influences on fixation durations. Empirical (a) and simulated data (b) contrasting fixation duration distributions in a memorization task compared to a visual search task. Direct comparison of simulated and empirical data for the (c) memory task and (d) for the search task.

Figure 6. Fixation duration as a function of scene onset delay. Each point represents one critical fixation, defined as the first fixation following the saccade in which the scene was removed from the display. Simulated data (a) vs. empirical data (b). The diagonal line represents critical fixations from the infinite delay condition in which the scene returned during the first saccade, i.e. scene onset delay = critical fixation duration.

Figure 7. Modeling fixation durations in the scene onset delay paradigm. From top to bottom, the sequence of events is organized along 4 time lines. (1) Scene disappearance time line: green horizontal line marks the absence of the scene. (2) Fixation duration time line: light-orange horizontal lines mark simulated fixation durations with actual duration on top; critical fixation duration in red; vertical broken lines visualize relation to saccade programming time line. (3) Random walk timing; onsets and offsets of single random walks are marked by a circle. Default transition rate r_1 in blue, reduced transition rate during scene absence, r_0, in green. Vertical broken blue lines visualize relation to saccade programming time line. Reaching threshold initiates a new labile saccade program. (4) Saccade programming time line: broken line segments represent labile saccade programs; solid red-colored line non-labile saccade programs; thick red bars represent saccade

duration (i.e., saccade execution); the green cross represents a saccade cancellations exerted in response to the disappearance of the scene.

Figure 8. (a) Probability of fixation duration prolongation as a function of scene onset delay. Simulated data (solid line) are compared with the empirical data (broken line). (b) Mean fixation durations for simulated and empirical data. Mean fixation duration in the 0 delay control condition is contrasted with mean critical fixation durations across the different delay conditions, separately calculated for the two observed populations of fixation durations. Linear regressions were fitted to the means across the scene onset delay conditions. In both panels, error bars represent the standard error of the mean.

Figure 9. Distribution functions of fixation duration in the scene onset delay experiment. Frequency of occurrence is calculated for each 60-ms bin. In a given panel, simulated data (solid line) are compared with empirical data (broken line). Vertical broken lines mark the delay duration. Note that in all (non-infinite) delay conditions, the second peak of the distribution rises about 100 ms following scene reinstatement.

Figure 10. Modeling fixation durations in the mask onset delay paradigm. Average fixation duration as a function of mask onset delay, for simulated (circles) and empirical data (squares). Error bars represent the standard error of the mean.

Figure 1

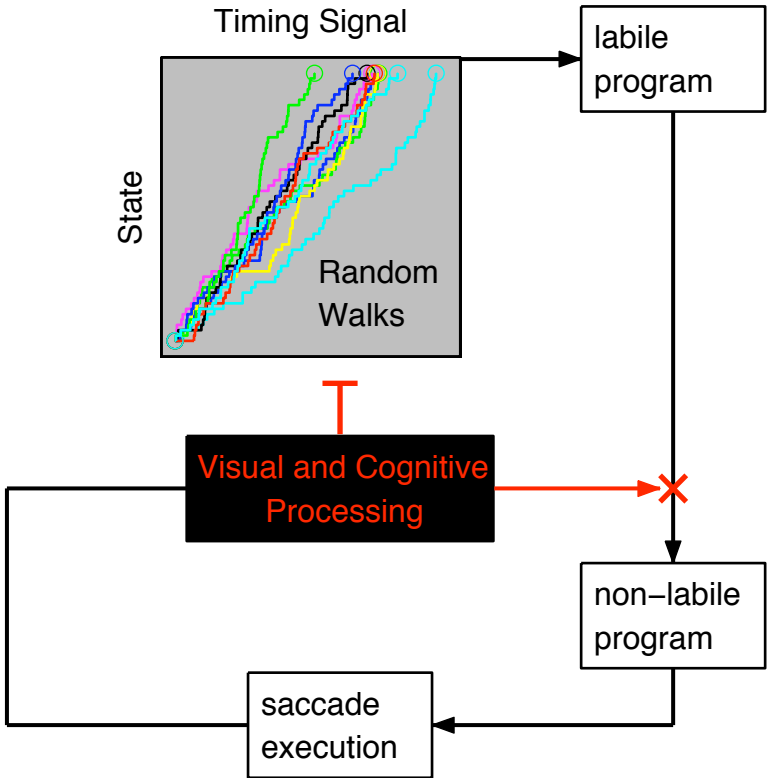

Timing Signal
State
Random
Walks
labile
program
Visual and Cognitive
Processing
non–labile
program
saccade
execution

Figure 2

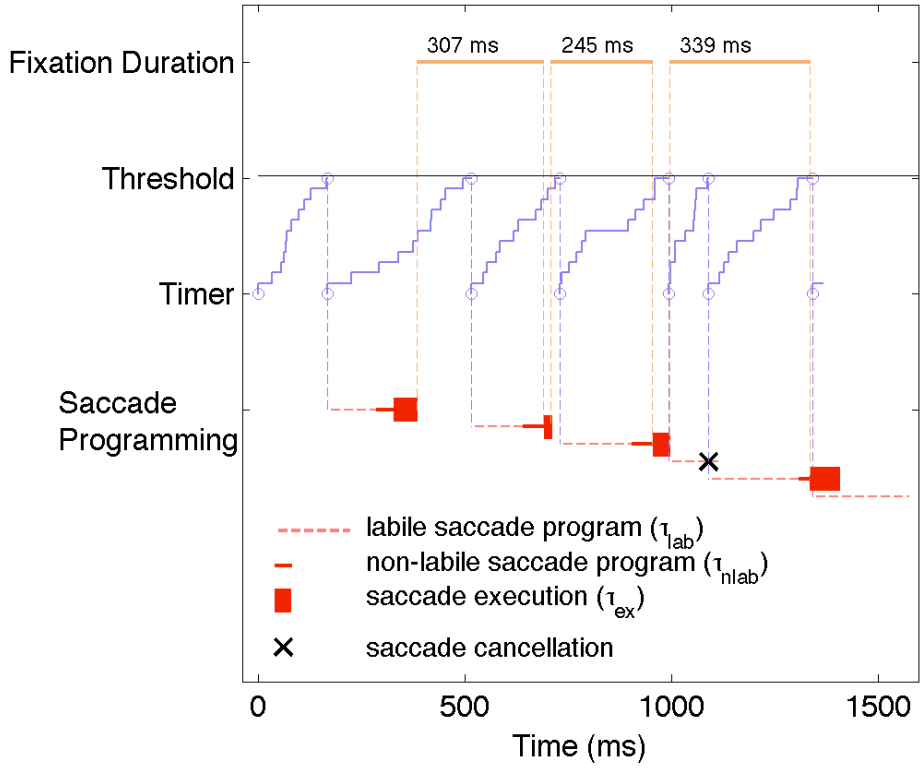

Fixation Duration
Threshold
Timer
Saccade Programming
307 ms
245 ms
339 ms
labile saccade program (τ_{lab})
non-labile saccade program (τ_{nlab})
saccade execution (τ_{ex})
saccade cancellation
0
500
1000
1500
Time (ms)

Figure 3

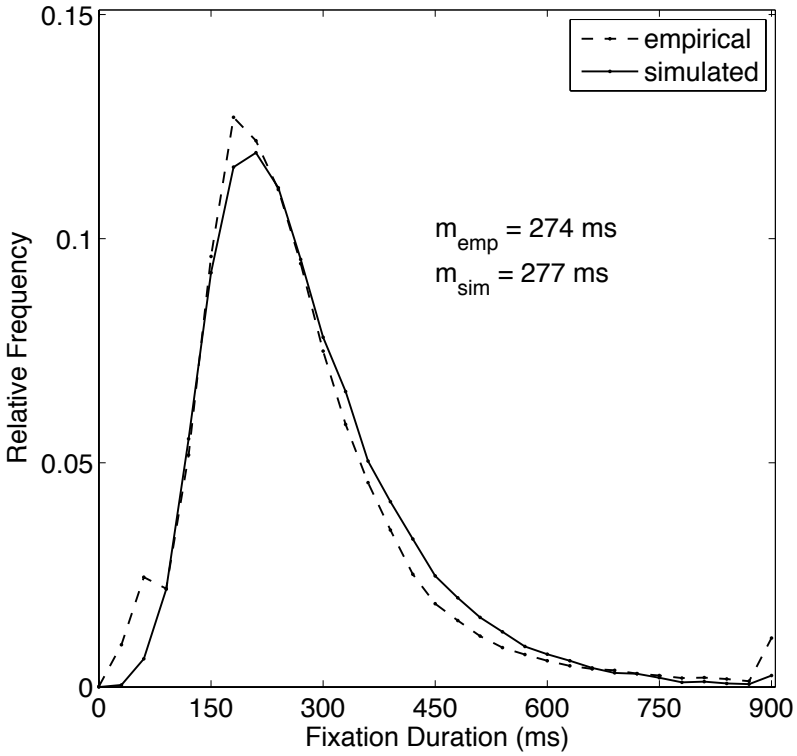

0.15
0.1
0.05
0
Relative Frequency
0
150
300
450
600
750
900
Fixation Duration (ms)
empirical
simulated
m_emp = 274 ms
m_sim = 277 ms

Figure 4

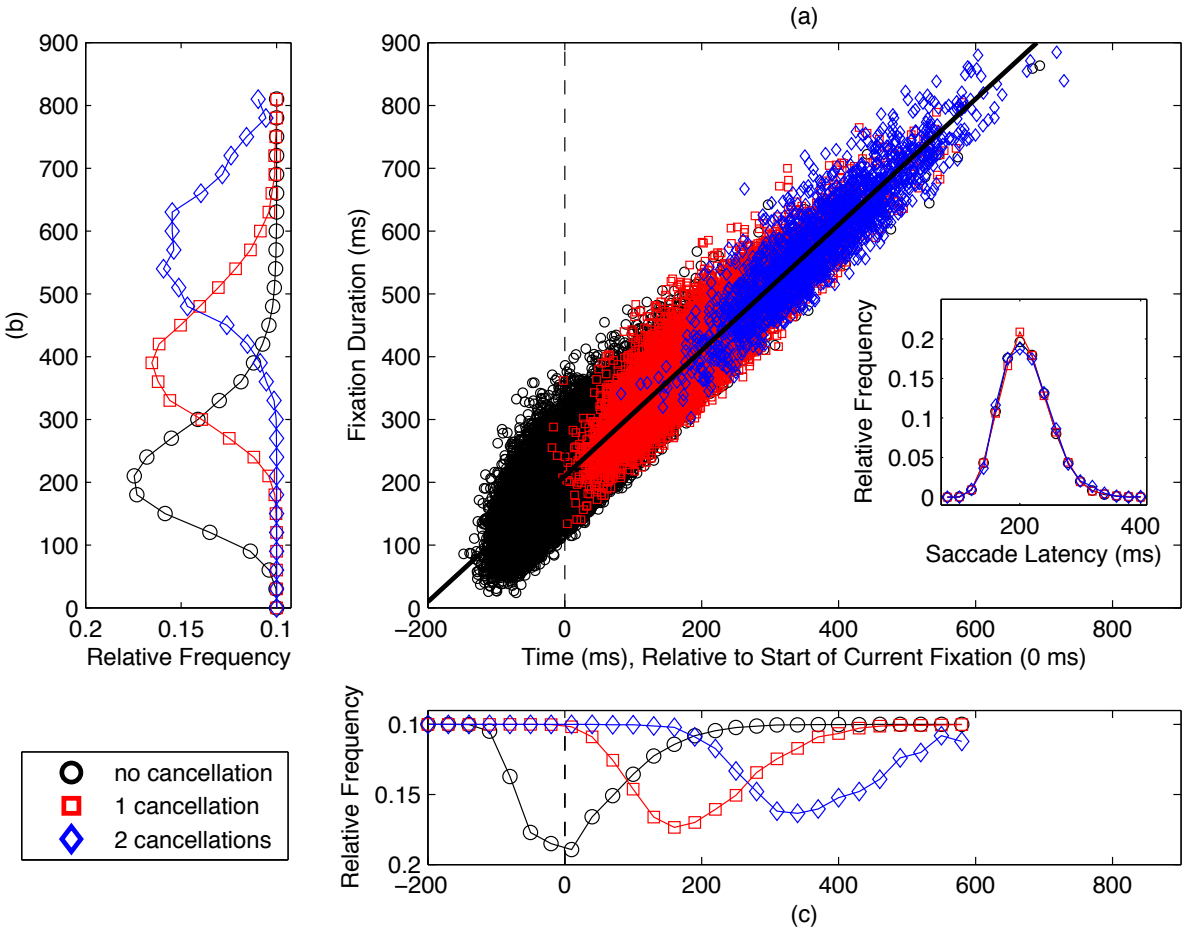

(a)
Fixation Duration (ms)
Time (ms), Relative to Start of Current Fixation (0 ms)
Relative Frequency
Saccade Latency (ms)
(b)
Relative Frequency
(c)
Relative Frequency
no cancellation
1 cancellation
2 cancellations

Figure 5

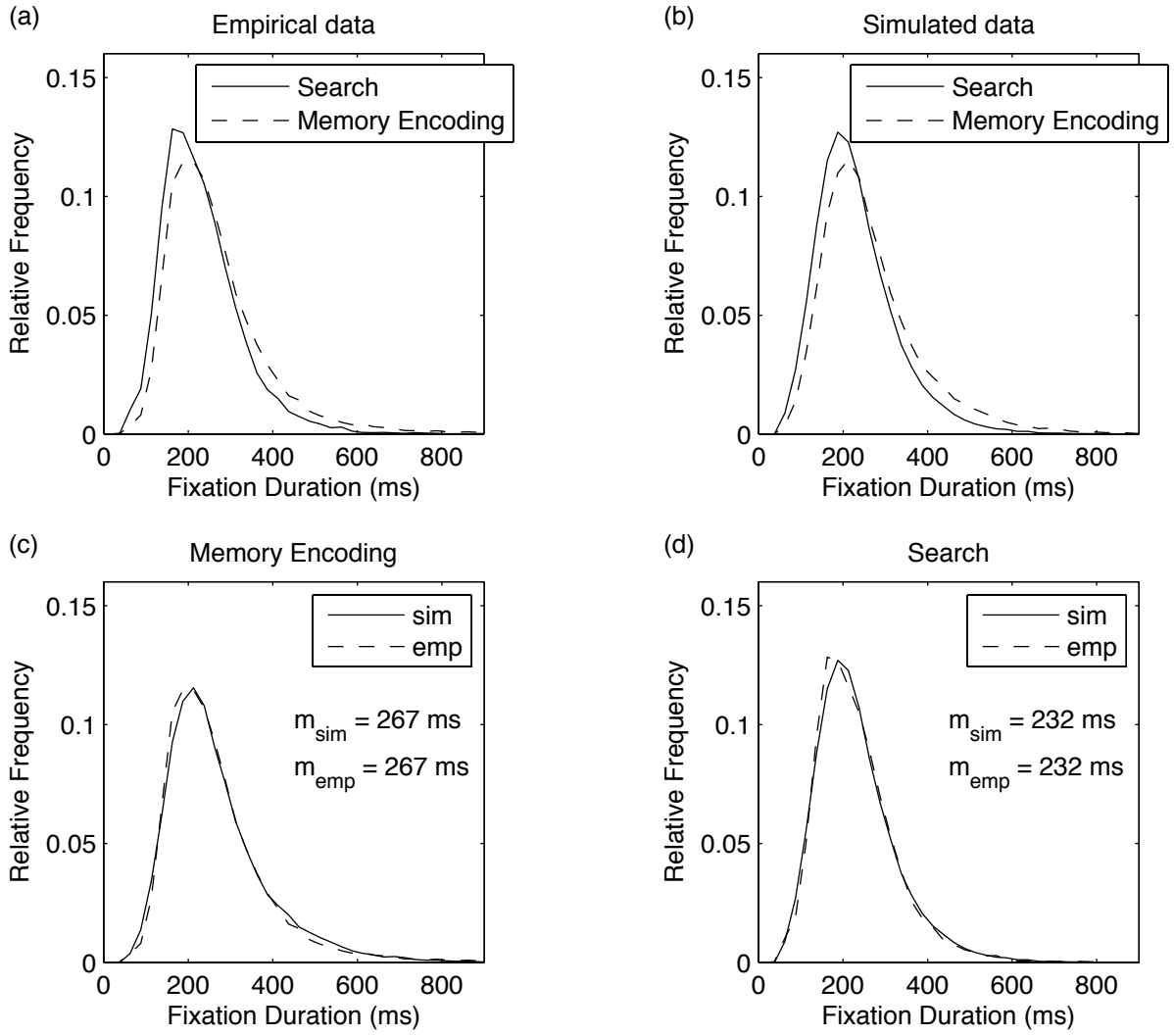

(a)
Empirical data
Search
Memory Encoding
Relative Frequency
Fixation Duration (ms)
(b)
Simulated data
Search
Memory Encoding
Relative Frequency
Fixation Duration (ms)
(c)
Memory Encoding
sim
emp
m_sim = 267 ms
m_emp = 267 ms
Relative Frequency
Fixation Duration (ms)
(d)
Search
sim
emp
m_sim = 232 ms
m_emp = 232 ms
Relative Frequency
Fixation Duration (ms)

Figure 6

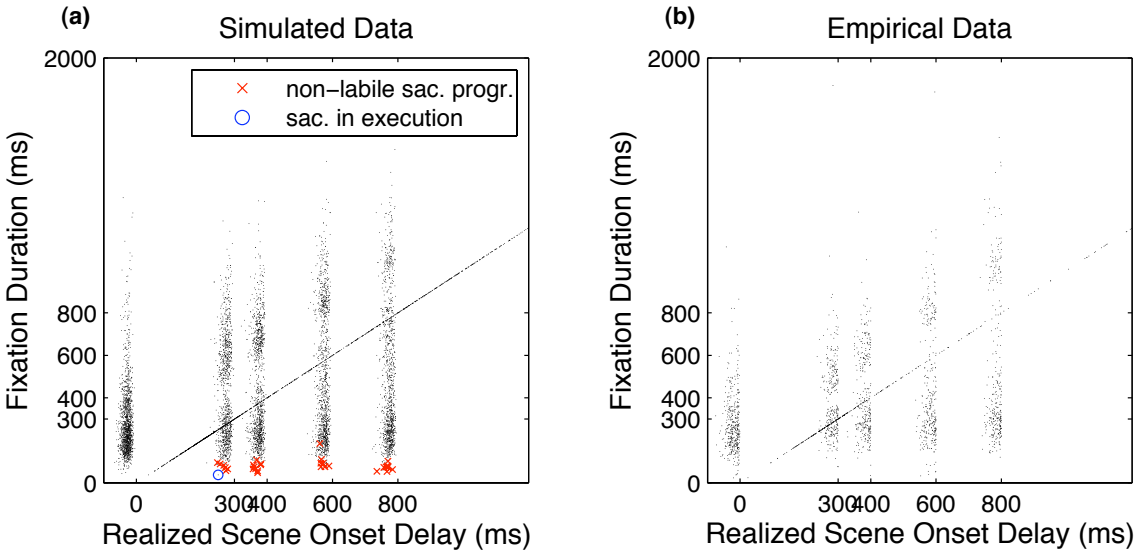

(a)
Simulated Data
non−labile sac. progr.
sac. in execution
Fixation Duration (ms)
Realized Scene Onset Delay (ms)
2000
800
600
400
300
0
0
300
400
600
800
(b)
Empirical Data
Fixation Duration (ms)
Realized Scene Onset Delay (ms)

Figure 7

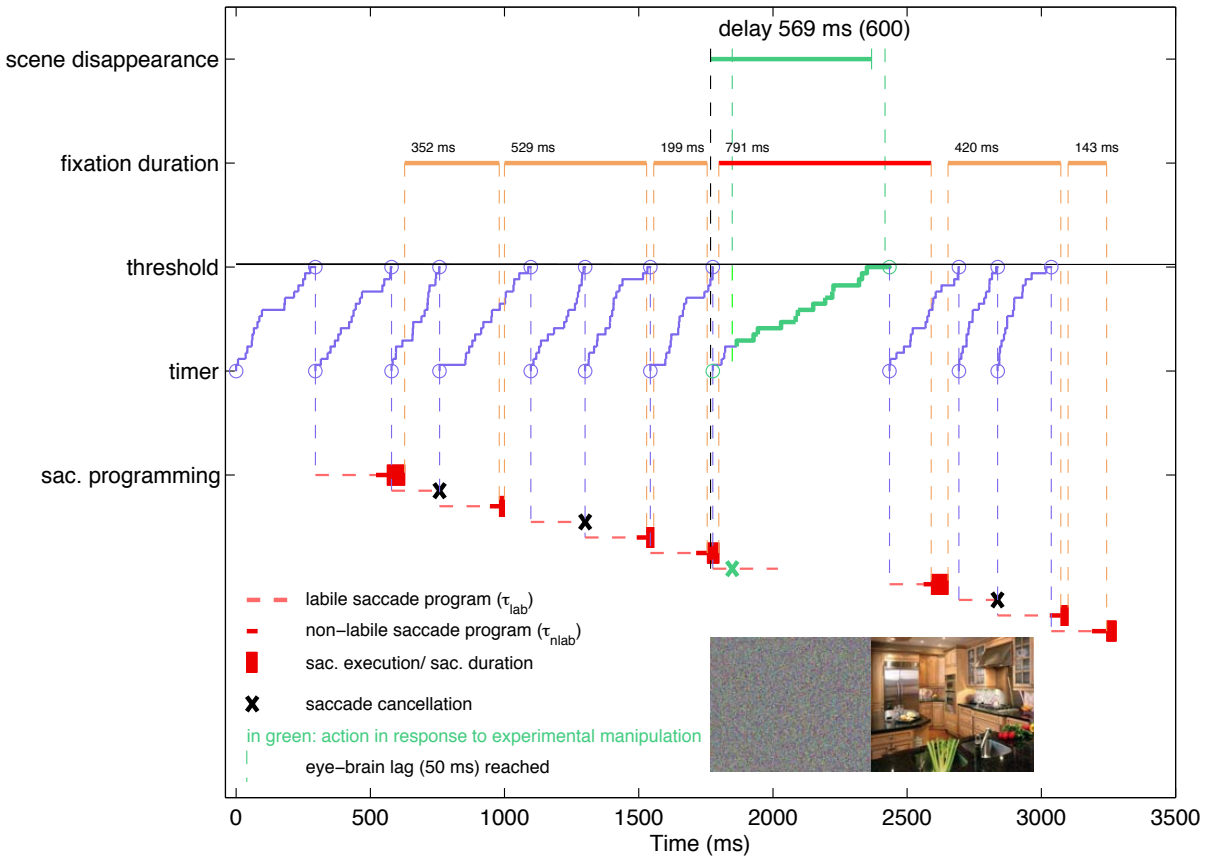

delay 569 ms (600)
scene disappearance
fixation duration
threshold
timer
sac. programming
352 ms
529 ms
199 ms
791 ms
420 ms
143 ms
labile saccade program (τlab)
non–labile saccade program (τnlab)
sac. execution/ sac. duration
saccade cancellation
in green: action in response to experimental manipulation
eye–brain lag (50 ms) reached
0
500
1000
1500
2000
2500
3000
3500
Time (ms)

Figure 8

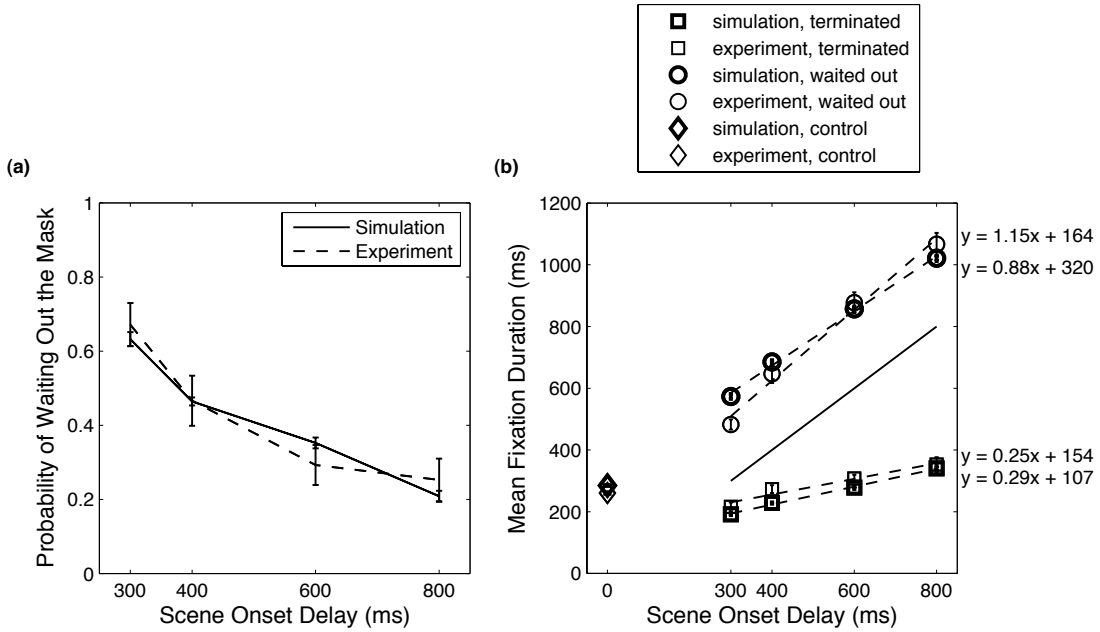

simulation, terminated
experiment, terminated
simulation, waited out
experiment, waited out
simulation, control
experiment, control
(a)
Probability of Waiting Out the Mask
Simulation
Experiment
Scene Onset Delay (ms)
(b)
Mean Fixation Duration (ms)
y = 1.15x + 164
y = 0.88x + 320
y = 0.25x + 154
y = 0.29x + 107
Scene Onset Delay (ms)

Figure 9

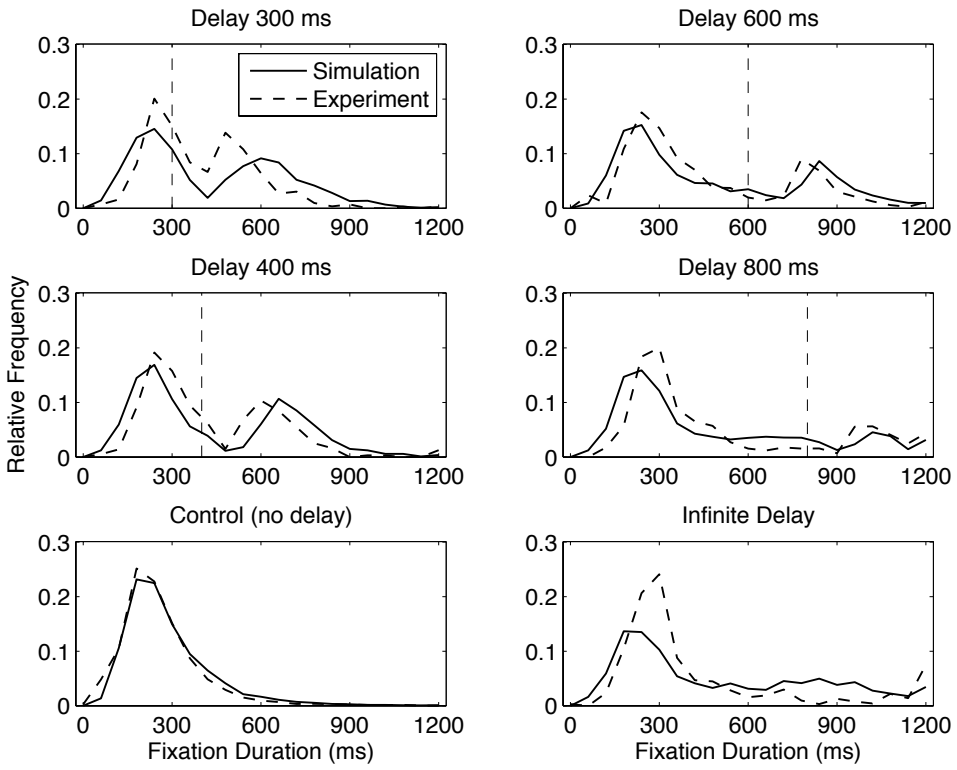
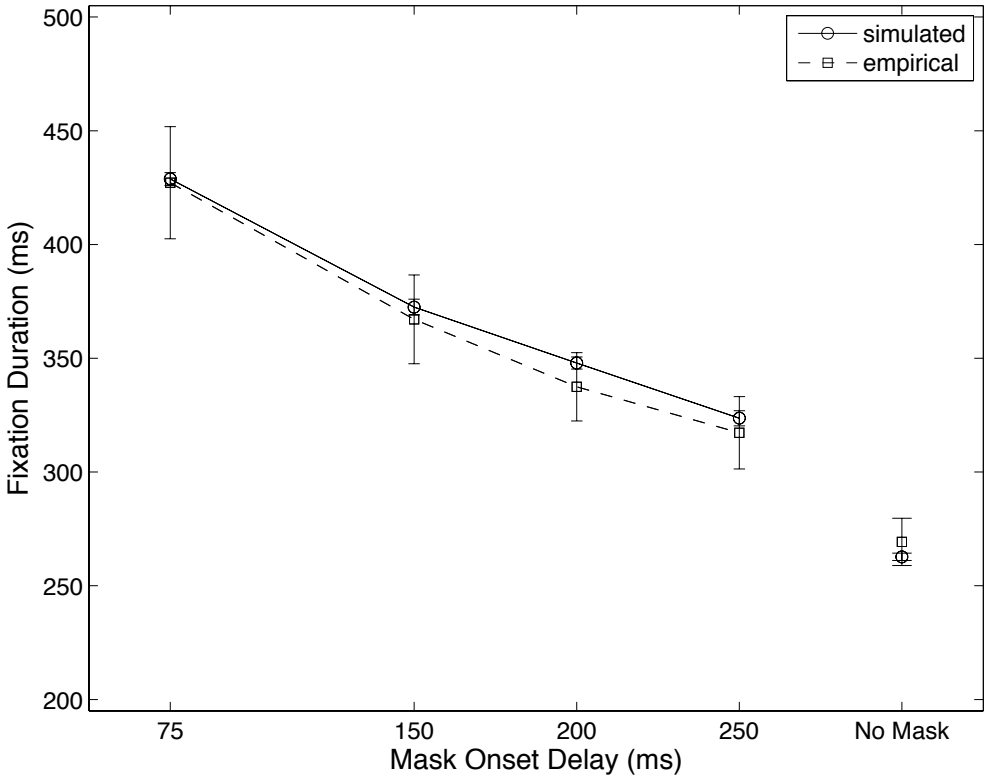

Delay 300 ms
Delay 600 ms
Delay 400 ms
Delay 800 ms
Control (no delay)
Infinite Delay
Simulation
Experiment
Relative Frequency
Fixation Duration (ms)
Fixation Duration (ms)
0
0.1
0.2
0.3
300
600
900
1200

Figure 10


simulated
empirical
Fixation Duration (ms)
500
450
400
350
300
250
200
75
150
200
250
No Mask
Mask Onset Delay (ms)